



Actes des
ATELIERS d' **INFORSID -**

*Dessínons ensemble le futur des
systèmes d'information*

INFORSID 2021
39^{ème} édition

Dijon, 1er au 4 juin 2021

Organisé en distanciel

Présidente du Comité de Programme: Selmin Nurcan

Présidents du Comité d'Organisation: Thierry Grison, Marinette Savonette

<https://inforsid2021.sciencesconf.org/>

Préface

J'ai le plaisir et l'honneur de présider le Comité de Programme d' INFORSID 2021.

Depuis 1982, le congrès annuel INFORSID (INFormatique des ORganisations et Systèmes d'Information et de Décision) constitue le lieu d'échange privilégié entre chercheurs et praticiens pour identifier et explorer les problématiques, les opportunités et les solutions que les systèmes d'information apportent ou absorbent. Les systèmes d'information sont omniprésents au sens où aucun domaine d'activité n'y échappe.

En 2021, le congrès INFORSID devait se tenir à Dijon du 1er au 4 juin 2021. La pandémie, que nous subissons depuis maintenant presque 18 mois au niveau mondial, en a décidé autrement. INFORSID 2021 s'est donc tenu pour la deuxième fois entièrement à distance.

INFORSID 2021 a accueilli trois ateliers cette année :

- *Évolution des SI: vers des SI Pervasifs ? (2ème édition)*, porté par Manuele Kirsch Pinheiro, Luiz Angelo Steffanel, Carine Souveyet et Philippe Roose
- *Systèmes d'Information et de Décision et démocratie dans les organisations (4ème édition)* porté par Raphaëlle Bour et Maryse Salles
- *SimpleText*, porté par Liana Ermakova, Josiane Mothe et Eric San-Juan

Les ateliers sont les pépinières des recherches de notre communauté. Je remercie le Bureau d'INFORSID et le Comité Exécutive d'avoir pleinement soutenu ma proposition de publier les articles acceptés dans les ateliers d'INFORSID, dans un deuxième volume, intitulé "*Actes des Ateliers d'INFORSID - Dessinons ensemble le futur des systèmes d'information*". J'espère que cela donnera plus d'élan aux bonnes volontés des prochaines années, pour organiser des ateliers et aussi pour soumettre des travaux à ces ateliers.

Par ailleurs, trois résumés étendus qui étaient initialement soumis pour la session internationale, ont été retenus par le Conseil du Comité de Programme pour publication dans les actes des ateliers et pour présentation dans une nouvelle session d'INFORSID, *le FORUM*, qui j'espère sera pérenne dans les prochaines éditions du Congrès.

Nous remercions chaleureusement

- les porteurs des ateliers pour leur investissement dans l'organisation de ces rencontres enrichissantes et originales (nos pépinières), qui participent à la fois au dynamisme de la conférence mais également à son renouvellement ;
- les auteurs d'avoir partagé leur récents travaux et leurs idées avec la communauté INFORSID ;
- les participants aux Ateliers et au Congrès INFORSID de faire vivre et promouvoir notre communauté.

Je tiens à remercier les membres du bureau de l'association INFORSID, sous la présidence de Franck Ravat, de m'avoir confié l'organisation scientifique du congrès et pour leur assistance et soutien tout au long de cette année.

Last but not least, je remercie toutes les personnes impliquées dans l'organisation d'INFORSID 2021, en particulier Thierry Grison et Marinette Savonnet, qui ont réalisé un travail essentiel pour la bonne tenue du Congrès sur une plateforme collaborative, Teams.

Nous sommes impatients de vous retrouver dans notre jumeau physique dans un avenir proche.

Très bonne conférence à tous,

Selmin NURCAN
Présidente du Comité de Programme INFORSID 2021

Sommaire

ATELIER Évolution des SI : vers des SI pervasifs ?

Atelier ” Evolution des SI : vers des SI Pervasifs ? ” - 2ème Édition	1
<i>Manuele Kirsch Pinheiro, Luiz Angelo Steffenel, Carine Souveyet and Philippe Roose</i>	
L’ingénierie des systèmes d’information pervasifs : Quelles sont les compétences attendues ?	3
<i>Carine Souveyet and Manuele Kirsch-Pinheiro</i>	
Retour d’expérience dans le développement d’un système d’information pervasif au service d’un territoire rural	13
<i>Thierry Antoine-Santoni, Bastien Poggi, Evelyne Vittori, Marielle Delhom, Ho Van Hieu, Alexandre Mameli, David Araujo, Jean-Marc Simeoni, Jean-Sébastien Gualtieri and Antoine Aiello</i>	
Défis des Systèmes d’Information pervasifs dans les bâtiments intelligent en modes BIM, BAM et BOOM !	21
<i>Christophe Ponsard and Mounir Touzani</i>	
Modeling Energy Consumption in SOA: Requirements and Current Status	31
<i>Jorge Andrés Larracochea, Philippe Roose, Sergio Ilarri, Sébastien Laborie, Yudith Cardinale and Mauricio Jacobo González</i>	
Enrichir les traces GPS des visiteurs à partir de données qualitatives : enjeux et propositions	40
<i>Melanie Mondo, Cécile Cayèré and Jérémy Richard</i>	
Le projet Pulse : vers la supervision des échanges dans un système IoT . .	49
<i>Jannik Laval, Boubou Thiam Niang, Imene Ghzaïel, Kenza Riahi, Baudouin Dafflon, Giacomo Kahn and Yacine Ouzrout</i>	

ATELIER Simple Text

Atelier SimpleText : Simplification et Vulgarisation des Textes Scientifiques	57
<i>Liana Ermakova, Josiane Mothe and Eric Sanjuan</i>	
Recherche de phrases parallèles à partir de corpus comparables pour la simplification de textes médicaux en français	61
<i>Remi Cardon and Natalia Grabar</i>	
Pour une démarche de communication multimodale de données scientifiques : de la recherche documentaire à l’infographie via le mind mapping	64
<i>Araújo Silvia and Hannachi Radia</i>	

Controlled language for text simplification: Concepts and implementation <i>Mike Unwalla</i>	75
Could automatic text simplification assist correction-revision of scientific texts written by non-native English speakers? <i>Helen Mccombe-Boudry</i>	80
Developing Simple Web Text for People with Intellectual Disabilities and to Train Artificial Intelligence <i>John Rochford</i>	88
 ATELIER SID et démocratie dans les organisations	
Atelier SID et démocratie dans les organisations <i>Raphaëlle Bour and Maryse Salles</i>	97
 FORUM	
EcoSoft : Un éco-label pour la durabilité des logiciels <i>Rébecca Deneckère and Gregoria Rubio</i>	99
Fouille de Processus de Configuration dans les Lignes de Produits <i>Camille Salinesi, Housseem Chemingui, Ines Gam, Raul Mazo and Henda Ben Ghezala</i>	103
Le Machine Learning pour un Context Mining Facility <i>Nourhène Nourhène Ben Rabah, Manuele Kirsch-Pinheiro, Benedicte Le Grand, Ali Jaffal and Carine Souwey</i>	107

Atelier

Evolution des SI : vers des SI Pervasifs ?

2^{ème} Édition

**Manuele Kirsch Pinheiro¹, Luiz Angelo Steffanel²,
Carine Souveyet¹, Philippe Roose³**

*1. Centre de Recherche en Informatique, Université Paris 1 Panthéon Sorbonne
90 rue Tolbiac, 75013 Paris, France*

Manuele.Kirsch-Pinheiro@univ-paris1.fr, Carine.Souveyet@univ-paris1.fr

*2. LICIS, Université de Reims Champagne Ardenne
Campus Moulin de la Housse, 51697 Reims Cedex2, France*

Luiz-Angelo.Steffanel@univ-reims.fr

*3. LIUPPA / IUT de Bayonne
IUT de Bayonne E2S UPPA, Université de Pau 64600, Anglet, France*

Philippe.Roose@iutbayonne.univ-pau.fr

MOTS-CLES. Systèmes d'Information, Systèmes d'Information Pervasifs, IoT, GreenIT, Architecture Orienté Services, SmartCities, SmartBuildings, Data Monitoring, trace numérique, trajectoire sémantique, ingénierie SI.

Préface

Les nouveaux fondements informatiques tels que le Cloud Computing, le Fog ou l'Edge Computing, l'IoT, les microservices, le Big Data, le Machine Learning, etc., mais également l'ensemble des usages en situation de mobilité que permettent ces nouvelles technologies poussent les entreprises à transformer leurs pratiques et leurs processus métiers. Les Systèmes d'Information deviennent désormais ouverts sur le monde. Ils ne sont plus limités aux seules frontières de l'organisation. Ils sont disponibles partout. Ils accompagnent les collaborateurs et les clients où qu'ils soient et à tout moment, s'intègrent dans l'environnement physique, interagissent avec d'autres objets, d'autres systèmes. On peut désormais parler de SI devenus pervasifs.

L'omniprésence de la technologie dans les environnements a des nombreuses implications sur les Systèmes d'Information, de leur infrastructure à leurs processus et activités métiers. Cette technologie est vue comme le moteur de l'innovation transformant en profondeur les organisations. Cette transformation implique de relever de nombreux défis comme, par exemple, le défi écologique, le défi sur les architectures et sur la gestion des infrastructures, celui de la gestion et de l'exploitation des données issues de ces infrastructures, et bien d'autres. L'objectif de cet atelier est ainsi de mieux comprendre cette nouvelle génération de SI, de comprendre les défis qu'elle pose et comment les nouvelles technologies et tendances dans l'informatique peuvent contribuer à cette évolution. A travers les contributions et les discussions suscitées par celles-ci, nous espérons pouvoir nous interroger sur cette nouvelle génération de SI : sommes-nous à l'aube d'un SI plus « intelligent » ?

Lors de cette 2^{ème} édition de l'atelier, nous avons le plaisir de compter sur six articles abordant différents aspects témoignant de l'évolution des SI. Les thèmes abordés sont particulièrement variés et riches : la consommation énergétique, la collecte et l'analyse des données IoT mais aussi l'exploitation des données GPS, le déploiement des bâtiments intelligents mais également le déploiement des SI Pervasifs sur des territoires ruraux, ainsi que les compétences nécessaires pour la maîtrise de cette nouvelle génération de SI.

A partir de propositions concrètes et intéressantes, ces articles suscitent la discussion et l'échange autour de cette nouvelle génération de SI et des tendances qui permettent aujourd'hui son apparition. A travers les discussions, nous souhaitons susciter une certaine synergie entre les différentes communautés appelées à être acteur de cette évolution.

Comité de Programme

- Amel Bouzeghoub, Télécom SudParis
- Bénédicte Le Grand, Université Paris 1 Panthéon Sorbonne
- Chantal Taconet, Télécom SudParis
- Carine Souveyet, Université Paris 1 Panthéon Sorbonne
- Frank Ravat, IRIT / Université Toulouse I Capitole
- Frédérique Laforest, LIRIS / INSA Lyon
- Luiz Angelo Steffemel, Université de Reims Champagne-Ardenne
- Manuele Kirsch Pinheiro, Université Paris 1 Panthéon Sorbonne
- Phillippe Roose, IUT de Bayonne/LIUPPA-T2I
- Rebecca Deneckere, Université Paris 1 Panthéon Sorbonne
- Sophie Chabridon, Télécom SudParis
- Thierry Antoine-Santoni, Università di Corsica Pasquale Paoli

L'ingénierie des systèmes d'information pervasifs : Quelles sont les compétences attendues ?

Carine Souveyet¹, Manuele Kirsch-Pinheiro¹,

*1. Centre de Recherche en Informatique, Université Paris 1 Panthéon-Sorbonne
90 rue de Tolbiac, 75013 Paris, France*

Carine.Souveyet@univ-paris1.fr, Manuele.Kirsch-Pinheiro@univparis1.fr,

RÉSUMÉ. Grâce à l'émergence de technologies tels que IoT, Mobile computing, Cloud computing, Fog/edge computing le Système d'Information (SI) d'une organisation peut intégrer des environnements informatiques intelligents, complexes, hétérogènes et mobiles au-delà de l'environnement physique de l'organisation afin d'être plus réactif et proactif dans les services et les processus métiers qu'il supporte. Cette nouvelle génération de SI, nommée Système d'Information Pervasif (SIP) nécessite d'adopter une vision très intégrée de l'ensemble des couches d'un SI qui va de la plus basse « architecture physique informatique » à la plus haute de « pilotage métier ». Une telle verticalité contraste avec les pratiques traditionnelles, dans lesquelles la complexité d'un SI est maîtrisée par un cloisonnement en couches et une gestion de celles-ci en « silo ». Les auteurs essaient de comprendre quelles sont les connaissances et les compétences attendues pour l'ingénierie de tels systèmes.

ABSTRACT . Thanks to the emergence of technologies such as IoT, Mobile computing, Cloud computing, Fog / Edge computing, the Information System (IS) of an organization can integrate intelligent, complex, heterogeneous and mobile computing environments beyond the physical environment of the organization in order to be more responsive and proactive in the services and business processes it supports. This new generation of IS, called Pervasive Information System (PIS) requires adopting a very integrated vision of all the layers of an IS, which goes from the lowest "physical IT architecture" to the highest "business management". Such verticality contrasts with "traditional" practices on IS, on which, the complexity of an IS is controlled by a partitioning into layers and a management of these in "silo". The authors try to understand what knowledge and skills are expected for the engineering of such systems.

Mots-clés : Systèmes d'Information Pervasifs, Ingénierie des SIP, Compétences.

Keywords: Pervasive Information Systems, SIP Engineering, Skills.

1. Introduction

L'émergence de nouvelles technologies dans les SI conduit à leur évolution vers une nouvelle génération de Systèmes d'Information, les Systèmes d'Information Pervasifs. Ces nouvelles technologies influencent, en premier lieu, les infrastructures utilisées par ces systèmes, mais leur influence ne se limite pas à ce niveau purement technique. Tous les niveaux d'un Système d'Information peuvent être impactés (Kirsch Pinheiro, 2021). Dans une vision schématique (pour ne pas dire simpliste), on peut considérer que ces nouvelles technologies et les opportunités qu'elles apportent sont susceptibles d'influencer les infrastructures, mais également les services offerts, les applications et processus métiers, voire même le pilotage de ces systèmes (voir Figure 1).

Les nouvelles technologies, comme l'IoT, le Cloud et Fog/Edge Computing, apportent plus de dynamisme aux Systèmes d'Information et permettent ainsi d'envisager des SI plus souples, capables de mieux s'adapter aux changements. La notion de contexte, qu'on peut définir comme toute information pouvant caractériser une action ou une entité (Dey, 2001), peut alors contribuer à atteindre cette souplesse, devenue nécessaire pour mieux tenir compte de l'environnement dynamique vers lequel se dirigent petit à petit les Systèmes d'Information. L'information de contexte peut ainsi être capturée et remontée, niveau par niveau, tels des événements, et contribuer à l'adaptation de chaque niveau et du système dans sa totalité.

La Figure 1 montre que les technologies positionnées à gauche comme le Cloud Computing, Edge Computing et le Machine Learning permettent d'influencer une ou plusieurs couches du SIP sans préciser lesquelles particulièrement. Les différents dispositifs IT et l'IoT (IoT & Device) sont positionnés en bas car ils ont un impact direct sur la couche infrastructure. Les éléments placés verticalement à droite sont les préoccupations transversales qui doivent être prise en compte dans l'ingénierie des SIP : la gestion du contexte, la sécurité et les aspects « green IT».

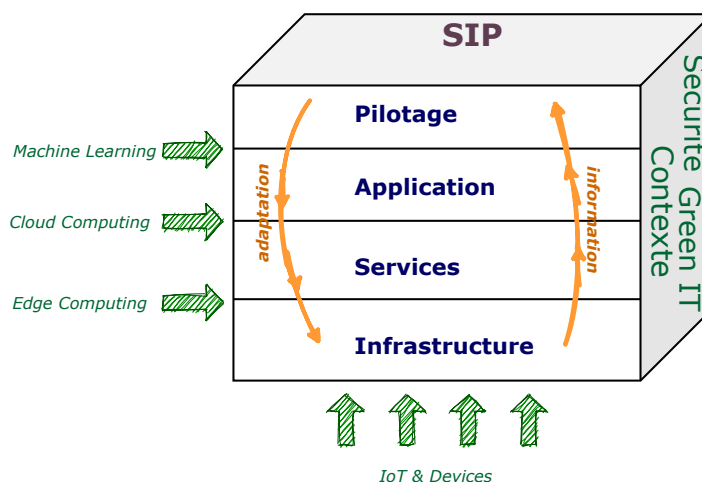


Figure 1. Vision schématique d'un Système d'information Pervasif

Grâce à ces nouvelles évolutions et pratiques, les SI peuvent s'étendre bien au-delà des limites physiques de l'organisation. Ils sont désormais accessibles partout, incluent des ressources internes et externes à l'organisation et peuvent même intégrer l'environnement physique. Les notions de ce qui est à l'intérieur ou à l'extérieur d'une entreprise sont devenues floues avec des processus utilisant des ressources autres que ceux du périmètre traditionnel de l'entreprise (Castro-Leon, 2014). L'environnement est devenu de plus en plus hétérogène, intégrant des dispositifs de nature très variée, qui peuvent en plus être mobiles, ce qui rajoute le dynamisme à l'hétérogénéité. Nous avons donc des SI qui sont de plus en plus confrontés à un environnement hétérogène et dynamique, intégrant des ressources de nature très variée (ressources Cloud, serveurs locaux, ordinateurs personnels, nano-ordinateurs, etc.) et des services tout aussi variés, pouvant appartenir à l'organisation elle-même, mais également à des organisations partenaires, voire même intégrer l'environnement physique entourant ces éléments et les utilisateurs. Il s'agit des systèmes pour lesquels on attend davantage de souplesse et une certaine « intelligence » afin de mieux réaliser les activités de l'organisation et les besoins de ses utilisateurs.

L'objectif de ce papier est de comprendre l'ingénierie nécessaire pour de tels systèmes et de définir quelles sont les compétences attendues dans l'ingénierie de ces Systèmes d'Information Pervasifs. Pour répondre à cette question, la section 2 précise la notion de « Systèmes d'Information Pervasifs » (SIP) considérée dans l'article ; les exigences attendues des SIP sont explicitées à la section 3. La discussion sur les compétences attendues pour cette nouvelle ingénierie est abordée à la section 4 en exploitant la verticalité du SIP et la convergence des disciplines autour de cette ingénierie, avant de conclure à la section 5.

2. Qu'est-ce qu'un système d'information pervasif ?

Plusieurs visions du terme « *Pervasive Information Systems* » existent. Une première tendance se résume par le fait que le mot clé « pervasif » est associé à l'information qui est omniprésente et qui est capturée n'importe où grâce à des capteurs dispersés dans l'environnement physique. Le système, dans ce cas, se conçoit comme un système orienté capteur pour capturer l'information n'importe où et n'importe quand. Cette tendance est représentée notamment par des systèmes issus de l'IoT, tels que (Xiao et al. 2017 ; Brahem et al. 2021 ; Lippi et al. 2021 ; Kim & Lee, 2021). Une autre tendance s'assimile plutôt aux environnements ubiquitaires, fournissant de manière autonome un confort à un ou plusieurs utilisateurs. Ces systèmes pervasifs sont souvent limités à une application, à un lieu et/ou à un ensemble de dispositifs technologiques intelligents mais sont rarement connectés aux métiers ou aux systèmes d'Information traditionnels d'une organisation. On parle alors de systèmes ou plutôt d'applications pervasives, dont on peut citer (Maass & Varshney, 2012 ; Cheraghi et al. 2021 ; Lalanda et al. 2021 ; Raychoudhury et al. 2013 ; Romero et al. 2010). Enfin, la tendance que nous considérons dans cet article est bien celle d'un Système d'Information qui devient pervasif, car il doit prendre en compte des événements qui se passent dans des environnements physiques et offrir

des services adaptés au plus près des utilisateurs. Dans cette tendance, nous pouvons citer (Kourouthanassis et al., 2010 ; Najjar et al. 2014 ; Hauser et al. 2017).

Lorsqu'on considère un Système d'Information qui devient pervasif, la synergie entre les couches métiers du Système d'Information et l'infrastructure IT distribuée, dynamique et hétérogène devient alors un facteur essentiel des SIP. Or, traditionnellement, l'ingénierie des Systèmes d'Information maîtrisait la complexité des SI en couches et en « silo ». Cette vision traditionnelle compartimentée limite les interactions entre les différents niveaux et ne favorise pas la circulation d'information entre eux. Or l'introduction des nouvelles technologies et des nouvelles pratiques a bouleversé cette organisation. Par exemple, le caractère stratégique du choix entre un déploiement « *on-premise* » ou sur le Cloud, ou encore la migration des systèmes vers une architecture micro-services sont autant d'illustrations de ces bouleversements. La complexité de gestion de l'infrastructure IT n'est plus qu'une question technique mais devient contrainte par des politiques qui sont dirigées par le métier. De la même manière, l'infrastructure IT omniprésente dans l'environnement physique d'une organisation permet de capturer de l'information qui peut avoir une influence sur les processus métiers supportés par le système d'information et de l'organisation. Cette synergie oblige à remettre en cause la strate de couches cloisonnées dans laquelle s'est construite l'ingénierie des SI et de penser le système d'information pervasif selon sa dimension verticale (verticalité) du métier à l'infrastructure IT de manière intégrée (voir la Figure 1).

3. Quelles sont les exigences attendues des Systèmes d'Information Pervasifs ?

Après avoir introduit notre vision du système d'information pervasif, nous allons discuter des exigences qui seront attendues par des SIP afin de mieux identifier les problèmes et enjeux à résoudre. En effet, différentes exigences ont été mise en avant dans (Najjar et al. 2014), dont certaines nous semblent particulièrement pertinentes en ce qui concerne l'ingénierie des Systèmes d'Information Pervasifs.

La première exigence attendue est la « *sensibilité au contexte* ». Déjà soulignée par (Kourouthanassis et al. 2010), il s'agit de la capacité d'un système à s'adapter à l'environnement qui l'entoure (Baldauf et al. 2007). Le SIP en tant qu'environnement ubiquitaire doit percevoir l'évolution de son environnement d'exécution et s'adapter en conséquence. La connaissance du contexte est utilisée pour obtenir cette capacité.

La deuxième est celle de « *la gestion de l'hétérogénéité* ». En effet, le SIP doit prendre en charge une variété très large de dispositifs techniques et de services proposés à l'utilisateur ou exécutés dans ces environnements physiques. La gestion de l'hétérogénéité est complexe mais primordiale.

La troisième exigence est relative à la « *transparence* » que l'on devrait retrouver dans l'Informatique Ubiquitaire. Comme l'a souligné Weiser (1991), l'Informatique Ubiquitaire devrait être invisible pour son utilisateur, disparaître dans l'environnement sans être distinguable de celui-ci. Le SIP en tant qu'environnement ubiquitaire doit cacher la complexité et l'hétérogénéité aux utilisateurs. Derrière cette notion de transparence se cache aussi la facilité d'utilisation d'un système dont la

manipulation ne doit pas demander d'effort particulier d'un point de vue cognitif, comme le souligne (Bell & Dourish, 2007).

La quatrième exigence concerne la « *satisfaction des exigences* ». Le SIP en tant que SI doit répondre aux besoins de l'utilisateur ou du métier n'importe où et n'importe quand. Il s'agit avant tout de rendre l'utilisateur plus productif, de proposer les moyens les plus adaptés pour satisfaire ses besoins, de manière à lui permettre de focaliser davantage sur les tâches génératrices de valeur ajoutée pour lui ou le métier.

La cinquième exigence attendue est « *l'adaptation* ». Le SIP doit pouvoir gérer la variabilité possible dans les services et les infrastructures techniques pour répondre aux besoins des utilisateurs et des métiers en toute circonstance. Il s'agit de pouvoir proposer un système qui puisse être perçu comme « intelligent » par sa capacité à s'adapter et à proposer le « bon service au bon moment ».

Les cinq exigences de base d'un SIP proviennent de la convergence entre celles relatives à un SI et celles associées aux environnements ubiquitaires ou pervasifs. La section suivante va se centrer sur l'ingénierie des SIP, et sa complexité de mise en œuvre par rapport à celle des SI traditionnels à cause de la *verticalité* qui a été introduite à la Section 2 et à la convergence des disciplines telles que IoT, périphériques intelligents, Cloud/Edge/Mobile Computing mais aussi le Machine Learning. En effet, la prise en compte de ces exigences engendre la verticalité mentionnée dans la Section 2, puisque chacun des niveaux illustrés dans la Figure 1 sera impacté à la fois par le respect de ces exigences en interne, mais également en réaction aux informations et aux adaptations issues/réalisées par les autres niveaux, et en influençant à son tour les autres niveaux.

4. Ingénierie des SIP : convergence des disciplines & verticalité

L'ingénierie des SIP consiste d'abord à expliquer les différentes compétences et connaissances attendues à chacune des couches de la Figure 1.

La couche « infrastructure » est de plus en plus complexe puisqu'elle intègre les technologies telles que Cloud/Edge et Mobile computing ainsi que IoT. Aujourd'hui pour maîtriser cette couche, il faut être capable de :

- gérer et adapter les infrastructures matérielles et logicielles utilisées traditionnellement dans les SI (matériel, plateforme d'exécution, réseau et middleware orienté service ou objet) ;
- gérer et adapter les infrastructures Cloud/Edge et Mobile computing et IoT pour étendre le SI vers les environnements pervasifs. Cette compétence nécessite des connaissances techniques sur l'hétérogénéité des dispositifs IT, leurs capacités mais aussi des connaissances sur les architectures micro-services et les techniques de virtualisation et les outils de déploiement associés ;

- Comprendre comment les politiques métiers peuvent/doivent impacter la gestion de l'infrastructure IT.

La couche « **services** » représente les services applicatifs (ou composants) qui sont déployés et exécutés sur l'architecture IT et qui permettent de supporter l'utilisateur et/ou le métier. L'orientation service est très connue au niveau des Systèmes d'Information et des applications. L'adoption d'une architecture micro-service remet au premier plan les architectures orientées services, pas au sens des technologies comme REST et SOAP, mais par rapport aux principes et aux qualités attendues par ces architectures, comme le souligne (Shadija et al. 2017). Les compétences et les connaissances utiles pour cette couche sont sur les architectures orientées services et les technologies existantes à base de service allant de la conception jusqu'à la partie déploiement et exécution sur l'architecture IT. Des middlewares et autres outils spécifiques sont utilisés pour généraliser et faciliter leur utilisation.

La couche « **application** » correspond à l'ensemble des applications constituant le Système d'Information. Chaque application est composée de services interconnectés entre eux par des règles applicatives. Le fonctionnement en silo dont ont bénéficié les SI pendant de longues années a contribué à la conception d'applications qui sont aujourd'hui considérées comme monolithiques et difficiles à faire évoluer. Le passage de celles-ci à une architecture micro-services est un sujet abordé par de nombreux travaux, tels que (Taibi & Systä, 2019 ; Da Silva et al., 2019 ; Balalai et al. 2015). Les compétences et connaissances requises sont relatives à la conception, la réalisation et la validation d'application selon une architecture orientée service. L'identification et la concrétisation de ces règles applicatives doivent être maîtrisées ainsi que celles de la variabilité du point de vue métier mais aussi du point de vue technique. La gestion de la variabilité et les possibilités d'adaptation, comme celles discutées dans (Baldauf et al. 2007 ; Lalanda et al. 2021 ; Raychoudhury et al. 2013 ; Romero et al. 2010), deviennent des connaissances majeures dans le contexte des SIP.

La couche « **pilotage** » permet d'avoir la vision globale des applications au niveau stratégique et au niveau métier. Traditionnellement, l'architecture d'entreprise, la gouvernance des SI et des processus métiers sont des thèmes de connaissances et de compétences utiles pour cette couche. Cette couche va devoir intégrer la valeur ajoutée des environnements pervasifs intégrées au niveau des couches : infrastructure IT, services et applications. A nouveau, l'exemple du passage à une architecture micro-service illustre bien l'impact de ces transformations, en principe techniques, sur l'architecture entreprise, comme le souligne, par exemple (Müssig et al. 2017).

Les couches pilotage, application, service et infrastructure doivent mettre tout en œuvre pour satisfaire les exigences des utilisateurs et du métier (quatrième exigence de la section 3).

La facette « **contexte** » mentionnée dans la Figure 1 est transversale aux autres couches. L'ingénierie de contexte, soulignée dans (Kirsch-Pinheiro & Souveyet, 2018), doit pouvoir s'appliquer à chacune des couches pour être capable de faire de l'adaptation. Les compétences et les connaissances nécessaires sont relatives à la

modélisation du contexte, l'acquisition du contexte, l'interprétation des données et aussi la prise en compte de la distribution du contexte. Cette ingénierie de contexte est primordiale dans les SIP puisque c'est cette fonctionnalité qui permettra de répondre aux exigences des SIP comme la sensibilité au contexte, la transparence et l'adaptation. L'architecture IT intégrera des dispositifs physiques ou logiciels (capteurs) pour acquérir les informations de contexte qui pourront ensuite être exploitées dans une ou plusieurs couches (flèche jaune « information » partant de « infrastructure » montant en traversant toutes les couches jusqu'au « pilotage » dans la Figure 1).

L'exploitation des informations de contexte venant des couches inférieures peut permettre à la couche concernée de réagir à la situation en changeant ou/et adaptant son propre fonctionnement, mais également la configuration de la couche inférieure (flèche jaune « adaptation » descendant de pilotage jusqu'à l'infrastructure).

Les informations de contexte capturées peuvent être utilisées par les techniques de ***d'Intelligence Artificielle et de Machine Learning*** pour prédire des comportements ou des situations indésirables qui pourraient être ainsi détectées et évitées. Même si la question de l'application de ces techniques à grande échelle (celle de tout un SI) persiste (Ben Rabah et al., 2020), il n'empêche que celles-ci offrent de nombreuses perspectives, comme l'illustrent des travaux tels que (Kim, & Lee, 2021 ; Lalanda et al. 2021 ; Rojas et al. 2020), faisant de ces techniques un pilier pour les SIP. Les compétences et connaissances sur les techniques dites de « Machine Learning » sont nécessaires pour mettre en œuvre l'adaptation désirée dans les SIP.

Tout aussi vertical que la notion de contexte elle-même, les SIP, de par leur intégration à l'environnement et leur caractère étendu, posent tout naturellement la question de la ***sécurité***, au sens large du terme, tant les implications dans ce domaine sont nombreuses. De la sécurité des informations à celle des services et des applications, du respect de la vie privée et autres interrogations nées de cette observation intensive du contexte, comme traité, par exemple, dans (Brahem et al. 2021), en passant par la notion de confiance aussi bien de l'utilisateur face au système, mais également des éléments de ces système entre eux, différents aspects sont concernés et tous les niveaux sont touchés de manière transversale. Les compétences autour de la sécurité se montrent ainsi de plus en plus nécessaires pour l'ingénierie des SIP.

Dans le même ordre idée, les préoccupations du cout énergétique des solutions adoptées tant du point de vue du matériel que du logiciel et de l'informatique durable ne doivent pas être exclues des SIP ainsi que son intégration dans un environnement plus écologiquement responsable (Sarkis, J., Koo, C. & Watson, R.T. (2013)). La facette « ***green*** » est aussi une préoccupation transversale ; c'est pour cette raison que les compétences autour de ***l'informatique durable (green)*** sont des compétences incontournables pour l'ingénierie des SIP.

Comme on peut le constater, la mise en œuvre des SIP nécessite des connaissances et des compétences dans plusieurs disciplines qui vont du pilotage du SI jusqu'aux infrastructures techniques et la maîtrise des nouvelles technologies au niveau infrastructure comme IoT, Cloud/Edge/Mobile Computing.

De plus, la synergie de toutes ces couches, par la boucle « information-adaptation » ne peut se faire que si, d'une part, l'ingénierie de contexte et les techniques de IA et Machine Learning sont maîtrisées et, d'autre part, par la verticalité des couches.

5. Conclusions

Les Systèmes d'Information Pervasifs doivent se voir comme une nouvelle classe de Système d'Information où l'on doit prendre en compte de manière fortement intégrée les environnements physiques outillés de dispositifs technologiques pour améliorer le métier, le support aux utilisateurs, et adapter au mieux la réponse du système d'information aux changements de l'environnement. Cet article montre que cette nouvelle classe remet en cause l'organisation en « silo » des couches décrivant les Systèmes d'Information. Aujourd'hui, les exigences en matière de SIP montrent que la synergie des couches est primordiale et que cette verticalité doit être explicite et complètement maîtrisée. Enfin, l'ingénierie des SIP implique la convergence de plusieurs domaines : SI, contexte, sécurité, informatique durable, IA et Machine Learning et la maîtrise de nouvelles technologies comme Cloud/Edge/Mobile Computing et IoT. On est donc face à des systèmes profondément pluridisciplinaires, où des compétences dans ces nombreux domaines s'avèrent nécessaires. Cette pluridisciplinarité, associée à la verticalité, interroge sur les modèles, méthodes et méthodologies nécessaires à l'ingénierie des SIP. Celle-ci, tout comme ces systèmes, sera profondément marquée par cette verticalité et cette pluridisciplinarité.

Bibliographie

- Baldauf, M., Dustdar, S., Rosenberg, F. (2007): A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing* 2(4), 263-277.
- Bell, G., Dourish, P. (2007): Yesterday's tomorrows: notes on ubiquitous computing's dominant vision. *Personal and Ubiquitous Computing*, 11 (2), Jan. 2007, 133-143.
- Ben Rabah, N., Kirsch Pinheiro, M., Le Grand, B., Jaffal, A., Souveyet, C. (2020): Machine Learning for a Context Mining Facility. 16th Workshop on Context and Activity Modeling and Recognition, 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 678-684.
- Brahem, M., Scerri, G., Anciaux, N., Issarny, V. (2021): Consent-driven data use in crowdsensing platforms: When data reuse meets privacy-preservation. 2021 IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 130-139.
- Castro-Leon, E. (2014): Consumerization in the IT service ecosystem. *IEEE IT Professional*, 16(5), September/October 2014, 20–27.
- Cheraghi, S., Namboodiri, V., Aarsal, G. (2021) : CityGuide: A Seamless Indoor-Outdoor Wayfinding System for People with Vision Impairments. *Mobile and Pervasive Assistive Technologies (MPAT 2021)*, 2021 IEEE Int. Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), 105-110.

- Dey, A. (2001): Understanding and using context. *Personal and Ubiquitous Computing* 5(1), 4-7.
- Hauser, M., Günther, S. A., Flath, C. M., Thiesse, F. (2017): Designing Pervasive Information Systems: A Fashion Retail Case Study. *Proceedings of the 38th International Conference on Information Systems (ICIS 2017)*.
- Kim, H., Lee, D. (2021): TAP: A Transformer based Activity Prediction Exploiting Temporal Relations in Collaborative Tasks. *17th Workshop on Context and Activity Modeling and Recognition (CoMoRea 2021), 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, 20-25.
- Kirsch-Pinheiro, M., Souveyet, C. (2018): Supporting context on software applications: a survey on context engineering (Le support applicatif à la notion de contexte : revue de la littérature en ingénierie de contexte). *Modélisation et utilisation du contexte*, 2(1), ISTE OpenScience. <https://www.openscience.fr/Le-support-applicatif-a-la-notion-de-contexte-revue-de-la-litterature-en/>.
- Kirsch Pinheiro, M. (2021) : Apports de la Notion de Contexte à Différents Systèmes. *Habilitation à Diriger de Recherches, Université Paris 1 Panthéon Sorbonne*. Disponible sur : <https://tel.archives-ouvertes.fr/tel-03122073> (dernière visite : avril 2021).
- Kourouthanassis, P. E., Giaglis, G. M., Karaiskos, D. C. (2010): Delineating 'pervasiveness' in pervasive information systems: a taxonomical framework and design implications. *Journal of Information Technology*, 25 (3), 273-287.
- Lalanda, P., Vega, G., Cervantes, H., Morand, D. (2021) : Architecture and pervasive platform for machine learning services in Industry 4.0. *PerCom Industry Track 2021, 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, 293-298.
- Lippi, M., Mariani, S., Zambonelli, F. (2021) : Developing a "Sense of Agency" in IoT Systems: Preliminary Experiments in a Smart Home Scenario. *17th Workshop on Context and Activity Modeling and Recognition (CoMoRea 2021), 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, 44-49.
- Maass, W., Varshney, U. (2012) : Design and evaluation of Ubiquitous Information Systems and use in healthcare. *Decision Support Systems*, 54(1), Dec. 2012, 597-609.
- Müssig, D., Stricker, R., Lässig, J., Heider, J. (2017): Highly Scalable Microservice-based Enterprise Architecture for Smart Ecosystems in Hybrid Cloud Environments. *Proceedings of the 19th International Conference on Enterprise Information Systems - Volume 1: ICEIS*, 454-459.
- Najar, S., Kirsch Pinheiro, M., Le Grand, B., Souveyet, C. (2014): A user-centric vision of service-oriented Pervasive Information Systems. *8th International Conference on Research Challenges in Information Science (RCIS 2014)*, IEEE, 359-370.
- Raychoudhury, V., Cao, J., Kumar, M., Zhang, D. (2013): Middleware for Pervasive Computing: A Survey. *Pervasive Mobile Computing*, 9 (2), 177-200.
- Rojas, J.A.R., Rosas, J., Shen, Y., Jin, H., Dey, A.K. (2020): Activity Recommendation: Optimizing Life in the Long Term. *IEEE International Conference on Pervasive Computing and Communications (PerCom 2020)*, 1-10. Doi: 10.1109/PerCom45495.2020.9127358.

- Romero, D., Rouvoy, R., Seinturier, L., Carton, P. (2010) : Service Discovery in Ubiquitous Feedback Control Loops. In: Eliassen, F., Kapitza, R. (Eds.), 10th IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems (DAIS) / International Federated Conference on Distributed Computing Techniques (DisCoTec), Lecture Notes in Computer Science, 6115, 112-125.
- Sarkis, J., Koo, C. & Watson, R.T. (2013) Green information systems & technologies – this generation and beyond: Introduction to the special issue. *Inf Syst Front* 15, 695–704 (2013). <https://doi.org/10.1007/s10796-013-9454-5>
- Shadija, D., Rezai, M., Hill, R. (2017): Towards an Understanding of Microservices. 2017 23rd International Conference on Automation and Computing (ICAC), 1-16. Doi 10.23919/ICoAC.2017.8082018.
- Weiser, M. (1991): The computer for the 21st century. *Scientific American*, 265(3), September 1991, 94-104.
- Xiao, B., Rahmani, R., Li, Y., Kanter, T. (2017): Edge-based interoperable service-driven information distribution for intelligent pervasive services. *Pervasive and Mobile Computing*, 40, Sept. 2017, 359–381.

Retour d'expérience dans le développement d'un système d'information pervasif au service d'un territoire rural

Thierry Antoine-Santoni¹, Bastien Poggi¹, Evelyne Vittori¹, Marielle Delhom¹, Ho Van Hieu¹, Alexandre Mameli, David Araujo, Jean-Marc Simeoni¹, Jean-Sébastien Gualtieri², Antoine Aiello³

1. Unité Mixte de Recherche Université de Corse – CNRS SPE 6134

*Quartier Grosseti, 20250 Corte, France
antoine-santoni@univ-corse.fr*

2. SITEC, Parc Technologique Erbajolo,

*20601 Furiani
jsgualtieri@sitec.corsica*

3. Unité Mixte de Service Université de Corse – CNRS Stella Mare 3514

*Lieu dit a marana, 20250 Biguglia, France
aiello@univ-corse.fr*

RESUME. Cet article présente le bilan dans le développement d'un système d'information au sein d'un territoire qui se veut intelligent. Cette expérimentation, d'une durée de trois ans, a permis de travailler sur la mise en place d'un système d'information pervasif basé sur un réseau de capteurs sans fil utilisant la technologie LoraWAN. Nous en évaluons les réussites mais nous énonçons un certain nombre de limites dans la réalisation.

ABSTRACT. This article presents the results of the development of an information system within a territory that aims to be intelligent. This three-year experimentation allowed us to work on the implementation of a pervasive information system based on a network of wireless sensors using LoraWAN technology. We evaluate the successes of this experiment but we also point out a certain number of limitations in its implementation.

KEYWORDS: pervasive information system, rural territory, experimentation

1. Introduction

De nos jours, les villes et les zones urbanisées se complexifient et nécessitent de plus en plus le recours à des dispositifs numériques dans l'environnement quotidien du citoyen afin d'alimenter des outils d'aide à la décision pour gérer la ville mais

également pour développer des services aux habitants. Cette volonté est au cœur des territoires intelligents. Un territoire intelligent peut se définir comme l'expression d'une vision et d'une ambition de développement d'un territoire, qui s'évalue en termes de bénéfices pour les habitants, en y intégrant des éléments socles : le développement durable et social, l'éducation, la protection de l'environnement, les services aux usagers.

Pour développer ces services et usages, le numérique est central. On retrouve naturellement la notion de système d'information à des fins de stockage, de visualisation et d'analyse des données mais également dans un objectif de coordination des services afin d'interagir avec les environnements. Ces interactions s'appuient souvent sur un ensemble d'entités d'observation distribué sur l'espace concerné, des réseaux de capteurs sans fil, qui convertissent le plus souvent des informations physiques en données numériques. Le coût sans cesse réduit des réseaux de capteurs sans fil permet de les intégrer partout pour surveiller et pour contrôler pratiquement n'importe quel espace et environnement. Ces équipements numériques vont ainsi transformer l'environnement physique en un système global cyber-physique capable de collecter des données, d'analyser et d'interagir et en rendant cet espace parfois même intelligent. Ces environnements nouveaux sont regroupés sous les termes d'informatique omniprésente, d'informatique ubiquitaire ou pervasive ou bien d'intelligence ambiante. Dans (Rocher et al. 2020)(Rocher 2020), les auteurs tentent de nous préciser ces différents concepts :

- L'informatique omniprésente (ubiquitous computing) : c'est le fait de déplacer des ressources computationnelles de l'ordinateur personnel vers des objets de la vie quotidienne (santé, smart building) ;
- Informatique pervasive est très fortement liée à la notion d'informatique omniprésente avec une spécialisation sur les intergiciels : approche architecturale de médiation et d'orchestration des services.
- Intelligence ambiante se situe plus comme une coopération entre les ressources informatiques embarquées sur les objets afin d'assister les personnes dans leur activité quotidienne, avec une intégration forte de la notion d'intelligence artificielle

Dans la littérature, les exemples d'application de ces concepts sont nombreux autour du service aux citoyens (Du et al. 2018) liés à l'environnement (eau, qualité de l'air (Yi et al. 2015), aux infrastructures et des services publics (bâtiments, routes, parking, production d'énergie) à l'agriculture (Tzounis et al. 2017) et l'élevage (Somov et al. 2018) mais également de la santé et du bien-être (Haque, Milstein, et Fei-Fei 2020). Dans (Antoine-Santoni, Poggi, Vittori, Araujo, et al. 2019), les premiers éléments constitutifs d'un système d'information pervasif étaient présentés afin de proposer une vision et une mise en application de ces nouveaux outils. Dans cet article, il est proposé dans un premier temps une description de l'architecture globale finale proposée sur le village avec ses éléments constitutifs. Dans une deuxième partie, nous abordons les réussites, les limites dans le déploiement et les vigilances sur la proposition de tel environnement au cœur de la vie des citoyens. Nous concluons enfin sur les perspectives de ce projet.

2. Le smart village

Nous présentons (Antoine-Santoni, Poggi, Vittori, Araujo, et al. 2019) les premières pistes de développement d'un système d'information contextualisé, tendant vers un système pervasif, dans le cadre du programme scientifique Smart Village-Smart Village : émergence de territoires intelligents. Ce système intègre des réseaux de capteurs sans fil afin de collecter de nombreuses informations. Ces dispositifs utilisent la technologie de communication LoRaWAN (Lora Alliance 2015) dans de nombreux secteurs (Antoine-Santoni, Poggi, Vittori, Manicacci, et al. 2019). La Figure 1 illustre les différentes applications du Smart Village intégrant les outils de production énergétique (chaufferie biomasse, microcentrales électriques) les bâtiments connectés, des stations météo, l'agriculture et l'élevage, la qualité de l'eau et de l'air.

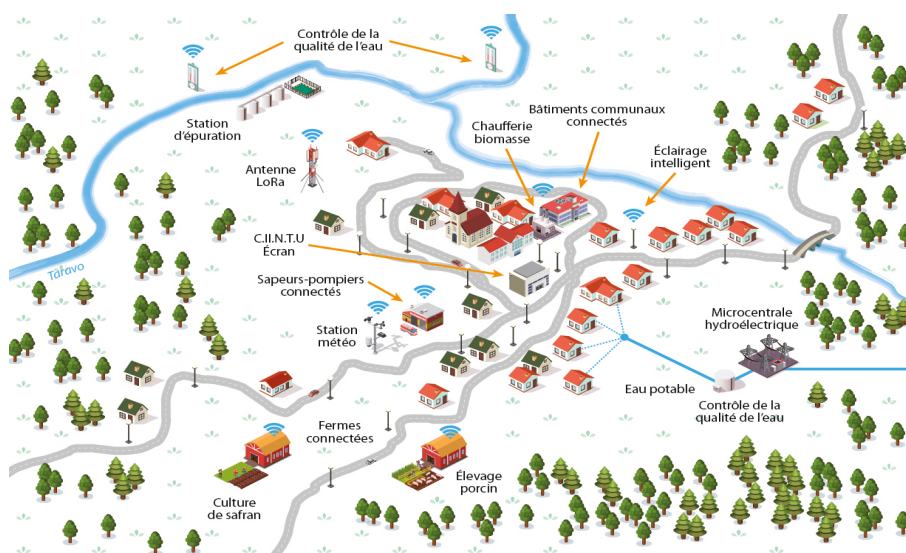


Figure 1 : Les applications du Smart Village

2.1 Le système d'information

L'intégralité des données produites est transmise dans un centre de données situé à Bastia auprès du partenaire privé du projet, la SITEC. Ces données sont ensuite traitées par des outils d'analyse et de visualisation des données. Ces outils se basent sur le moteur Elastic et son interface associée Kibana. Ils permettent entre autres, une bonne interactivité avec l'utilisateur (sélection possible directement dans les graphiques, modification de la présentation des graphiques, des tris, etc...), des formats de restitution divers et variés (histogrammes, courbes, secteurs, aires, données géolocalisées...) et sont optimisés pour les écrans tactiles. L'architecture mise en place pour le recueil, le stockage et la restitution des données issues de capteurs repose sur un cluster de quatre serveurs dédiés : trois pour le stockage et l'analyse et un pour

la supervision comme montré sur la Figure 2. L'interface de visualisation des données permet trois types de visualisations avec des tableaux de bord associés :

- une interface permettant de faire des extractions brutes, larges et accessibles rapidement pour les scientifiques ;
- une interface dédiée au maire et aux acteurs économiques (éleveurs, agriculteurs, etc.) avec des données filtrées, traitées ;
- une interface pour les habitants permettant d'afficher des données lisibles, compréhensibles et validées.

Les données recueillies proviennent de sources hétérogènes et possèdent des formats différents. Le système dispose aussi bien de données issues des 17 groupes de capteurs déployés sur le village de Cozzano, de données provenant de suivi de déplacement GPS d'animaux, de données de production et de consommation électrique (provenant soit d'EDF à l'aide de compteurs de consommation agrégée comme le montre la Figure 3, soit de la microcentrale hydraulique, soit de la chaudière biomasse ou de panneaux solaires).

Plusieurs contraintes ont été prises en compte lors de la conception de cet outil de supervision :

- l'authentification et la sécurisation des accès aux données,
- l'affichage des tableaux de bord. L'intégralité des données indexées en base de données sont restituées sous formes de représentation graphique. De plus, des interfaces dédiées ont été développées afin de répondre au mieux aux attentes des chercheurs, du safranier illustré par la Figure 4 et de l'éleveur porcin comme montré sur la Figure 5.
- la compatibilité d'affichage sur différents supports comme les PC, tablettes et smartphones dans un développement en responsive design.

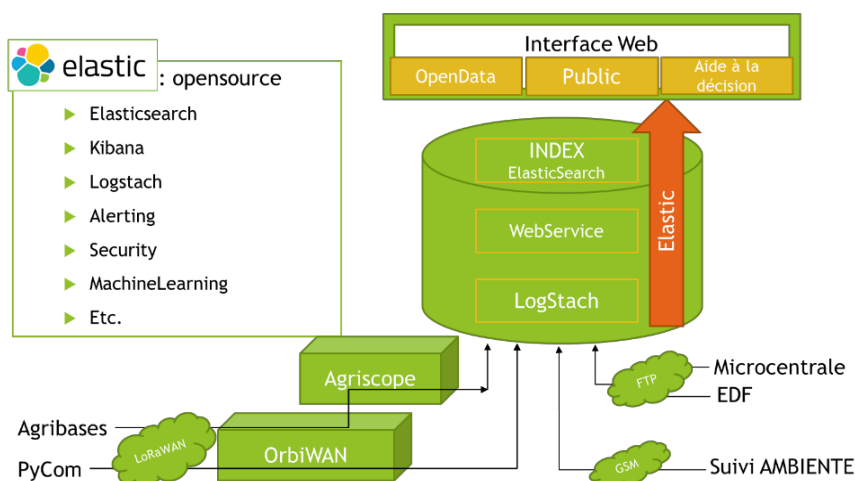


Figure 2 : Architecture des serveurs de données

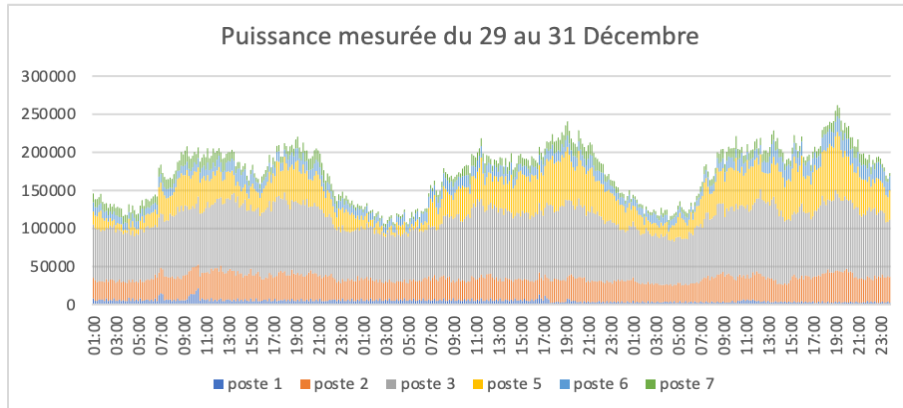


Figure 3 : Données agrégées de consommation en fin d'année 2020

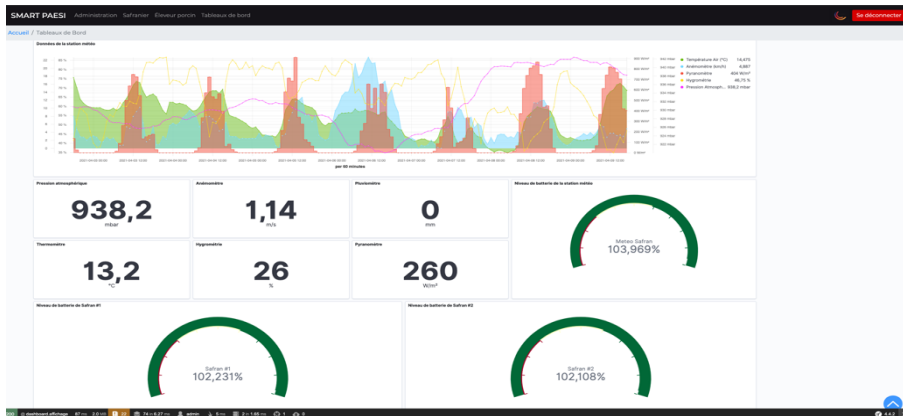


Figure 4 : Tableau de bord pour le suivi de l'exploitation de safran

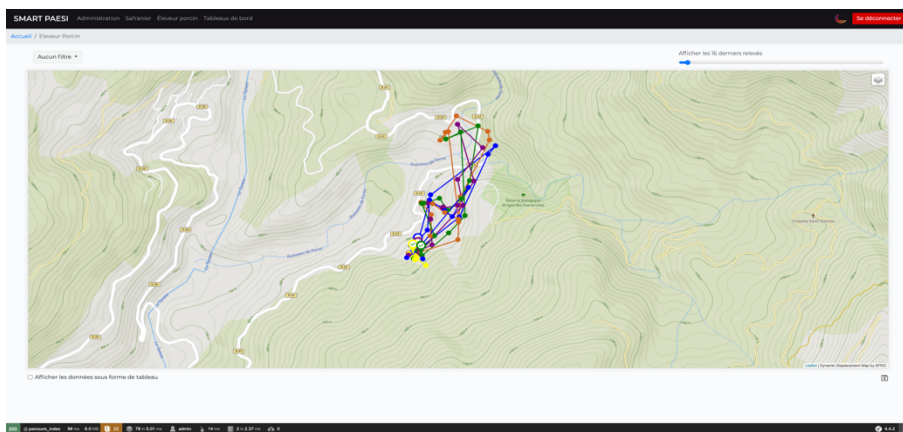


Figure 5 : Tableau de bord pour le suivi des troupeaux

2.2 Intégration de la pervasivité au sein du système d'information

Afin de rendre le système intelligent et en capacité de prendre des décisions en fonction du contexte, des données collectées sur le système observé sont traitées par une bibliothèque de fonctions d'apprentissage et d'optimisation intégrée dans un service web dans le but d'apprendre et de produire un ensemble de résultats de simulation. Ces résultats (R) sont produits à partir de données environnementales collectées par les réseaux de capteurs (E), des décisions prises ou à prendre (D). Cette bibliothèque de fonctions appelée Smart System (Antoine-Santoni, Poggi, Vittori, Hieux, et al. 2019) est décrite sur la Figure 6 à l'aide du formalisme DEVS (Zeigler, Prahofer, et Kim 2000).

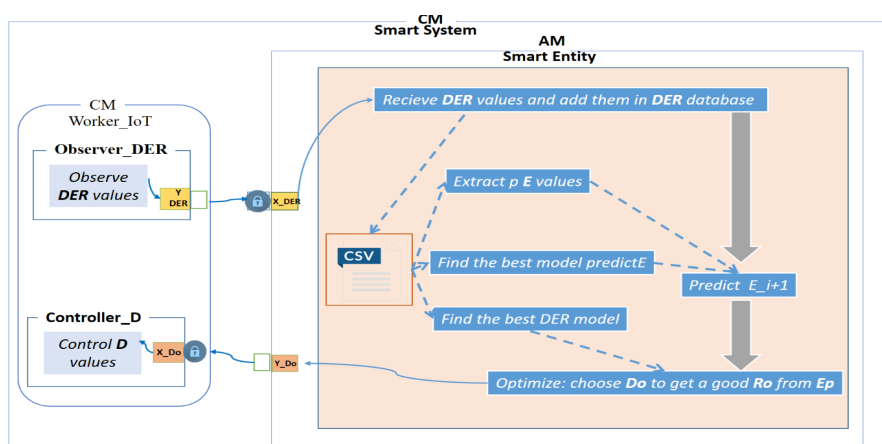


Figure 6 : Smart System, librairie d'algorithmes d'apprentissage sur des séries temporelles à des fins prédictives et d'interactions avec les réseaux de capteurs

Le *Smart System* est en mesure de réaliser des prédictions sur l'évolution du système observé par les réseaux de capteurs et d'interagir avec le réseau de capteurs afin de proposer une réponse optimisée et adaptée à l'évolution souhaitée du système. Les résultats de simulation sont probants mais ils n'ont pas encore été déployés dans le cadre du système réel. Il est nécessaire de réaliser de nombreuses simulations et tests avant de pouvoir passer à la mise en service.

3 Conclusion et perspectives

Le bilan de cette expérimentation met en lumière un certain nombre de réussites mais également de nombreuses limites dans la réalisation. Concernant les aspects positifs, nous pouvons citer :

- le déploiement du réseau sans fil LoRaWAN qui a rendu robuste la partie réseaux de capteurs sans fil du système d'information en favorisant la contrôlabilité, la rareté des défaillances et la faible maintenance ;

- la collecte, le stockage, l'utilisation et la visualisation de données ont été bien pensés permettant une visualisation dynamique et pertinente des données ;
- le Smart System et ses résultats de laboratoire, nous permettent de fonder de bons espoirs dans la mise en œuvre d'une réponse optimisée des dispositifs déployés avec l'environnement et les phénomènes observés par le SI. La place de la décision de l'homme pourra être minorée dans de nombreux cas.
- Le travail conceptuel du SI avec les futurs utilisateurs du village ; SI construit et pensé pour et avec eux.

Les limites s'expriment à plusieurs niveaux dans notre travail.

Le premier point est l'atteinte de la notion de la pervasivité de notre système à la fin de cette première phase d'expérimentation. En effet, la prise en compte du contexte dans la production et l'affichage des données a été un axe permanent dans notre développement mais cela se limite aux capacités des matériels et des technologies déployés. Une intelligence distribuée au sein du réseau de capteurs sans fil permettrait de franchir un seuil dans une véritable intelligence ambiante.

Le périmètre d'action d'un système d'information pervasif est une limite observée. En effet, les milieux clos tels que bâtiments sont le lieu d'application évident des systèmes d'information pervasifs. Les espaces naturels offrent naturellement des limites dans l'équipement technologique. Il existe cependant des opportunités, en particulier dans le domaine de la géolocalisation.

L'acceptation sociétale de tels dispositifs s'annonce comme un réel défi. Un travail important de sensibilisation et d'explication a permis de réduire les craintes mais le citoyen semble rapidement stopper la possible intrusion de tels dispositifs s'il n'en comprend pas clairement l'utilité dans sa vie privée, comme dans sa vie professionnelle. Cela est un frein important dans le développement des usages.

La sécurité est, naturellement, un élément crucial qu'il est nécessaire de traiter très amont – la multiplication des objets connectés, au sein des environnements, crée des possibles failles – la notion de canal auxiliaire (*side channel*) nous est clairement apparue durant cette expérimentation.

Le Smart Village devrait devenir un laboratoire vivant pour le développement d'une intelligence ambiante, coordonnée par sur un système d'information, au service d'un développement intelligent des territoires ruraux. Il permettra de continuer à formaliser les concepts, à développer de nombreux systèmes et dispositifs connectés, d'expérimenter des usages et d'en favoriser leurs acceptations, en lien permanent avec la population.

Bibliographie

- Antoine-Santoni, Thierry, Bastien Poggi, Evelyne Vittori, David Araujo, Ho Van Hieu, et Antoine Aiello. 2019. « Vers un système d'information pervasif pour un Smart Village ». in *Vers des SI pervasifs*. Paris.
- Antoine-Santoni, Thierry, Bastien Poggi, Evelyne Vittori, Ho Van Hieux, Marielle Delhom, et Antoine Aiello. 2019. « Smart Entity – How to Build DEVS Models from Large Amount of Data and Small Amount of Knowledge ? » P. 615-26 in *Proceedings of 11th EAI International Conference on Simulation Tools and Techniques*. Chengdu, People's Republic of China: EAI.
- Antoine-Santoni, Thierry, Bastien Poggi, Evelyne Vittori, François-Marie Manicacci, Jean-Sebastien Gualtieri, et Antoine Aiello. 2019. « Proposition of a smart environment architecture for resources monitoring and rural activities management ». P. 62-68 in *Proceedings of SENSORCOMM 2019*. Nice, France.
- Du, Rong, Paolo Santi, Ming Xiao, Athanasios V. Vasilakos, et Carlo Fischione. 2018. « The Sensable City: A Survey on the Deployment and Management for Smart City Monitoring ». *IEEE Communications Surveys & Tutorials* 1-1. doi: 10.1109/COMST.2018.2881008.
- Haque, Albert, Arnold Milstein, et Li Fei-Fei. 2020. « Illuminating the Dark Spaces of Healthcare with Ambient Intelligence ». *Nature* 585(7824):193-202. doi: 10.1038/s41586-020-2669-y.
- Lora Alliance. 2015. *White Paper: A Technical Overview of Lora and Lorawan*. CA, USA.
- Rocher, Gerald. 2020. *Thèse de doctorat : Évaluation de l'effectivité des systèmes ambiants*. Nice.
- Rocher, Gérald, Jean-Yves Tigli, Stéphane Laviolette, et Nhan Le Thanh. 2020. « Overview and Challenges of Ambient Systems, Towards a Constructivist Approach to Their Modelling ». *ArXiv:2001.09770 [Cs]*.
- Somov, Andrey, Dmitry Shadrin, Iliia Fastovets, Artyom Nikitin, Sergey Matveev, Ivan seledets, et Oleksii Hrinchuk. 2018. « Pervasive Agriculture: IoT-Enabled Greenhouse for Plant Growth Control ». *IEEE Pervasive Computing* 17(4):65-75. doi: 10.1109/MPRV.2018.2873849.
- Tzounis, Antonis, Nikolaos Katsoulas, Thomas Bartzanas, et Constantinos Kittas. 2017. « Internet of Things in Agriculture, Recent Advances and Future Challenges ». *Biosystems Engineering* 164:31-48. doi: 10.1016/j.biosystemseng.2017.09.007.
- Yi, Wei, Kin Lo, Terrence Mak, Kwong Leung, Yee Leung, et Mei Meng. 2015. « A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems ». *Sensors* 15(12):31392-427. doi: 10.3390/s151229859.
- Zeigler, Bernard P., Herbert Praehofer, et Tag Kim. 2000. *Theory of Modeling and Simulation 2nd Edition*. USA: Academic Press.

Analyse SWOT des Systèmes d'Information pervasifs dans les bâtiments intelligents sur les modes BIM, BAM et BOOM !

Christophe Ponsard, Mounir Touzani

1. CETIC - Centre de recherche, Gosselies, Belgique

christophe.ponsard@cetic.be

2. Chercheur indépendant, Toulouse, France

mounir.touzani@inrae.fr

RÉSUMÉ. Le concept de bâtiment intelligent date du siècle dernier et a été depuis longtemps soutenu par des systèmes d'information (SI). Cependant, ces SI sont devenus de plus en plus pervasifs. Et ce, suite aux effets conjoints de la disponibilité à coût abordable de réseaux de communications haut débit et sans fil, de capteurs formant un Internet des Objets et de capacités de traitements de grandes quantités d'informations à la demande. Cet article vise à alimenter l'atelier dédié à cette thématique en repartant de la définition du concept assez mal défini de bâtiment intelligent. Sur cette base, il examine les types de SI qui s'y sont déployés, identifie les parties prenantes concernées et une série d'exigences communes. Une analyse SWOT est ensuite esquissée afin d'identifier notamment certaines problématiques importantes à prendre en considération pour déployer et faire évoluer des SI dans ce contexte. Nos propos sont soutenus par une analyse sur un déploiement en cours au sein du bâtiment de notre centre de recherche et de la littérature.

ABSTRACT. The concept of Smart Building dates back to the last century. It has been supported by Information Systems (IS) for long. However, these IS are becoming ever ubiquitous as a result of the joint effects of the cheap and widely availability high-speed connections, the emergence of the Internet of Things and of Big Data processing capabilities. This article aims at providing the workshop by looking back at the Smart Building definition. On this basis, it examines the kind of IS deployed, it identifies relevant stakeholders and a series of common requirements. A SWOT analysis is then sketched in order to identify important issues to be considered in order to deploy and develop IS in this context. Our analysis also relies on an ongoing deployment of such a system within the building of our research center and some literature reports.

MOTS-CLÉS : Bâtiment intelligent, efficacité énergétique, transition numérique, systèmes d'information pervasifs

KEYWORDS: Smart Building, Energy Efficiency, digital transformation, Ubiquitous Information Systems

1. Introduction

Depuis une dizaine d'années, la transformation numérique est à l'oeuvre dans de nombreux domaines de notre quotidien. Les avancées technologiques liées à l'explosion de notre capacité à collecter, stocker, traiter et diffuser de l'information, a d'abord permis de soutenir les processus en place puis d'innover en remodelant notre manière de collaborer, de nous structurer et de conduire nos métiers.

Le domaine de la construction n'a pas échappé à cette évolution avec l'émergence du concept de bâtiment intelligent (ou « Smart Building »). Cette intelligence est plus à mettre en relation avec l'efficacité avec laquelle les systèmes d'information (SI) impliqués permettent aux utilisateurs de réaliser leurs tâches. La définition précise de ce terme reste difficile à décrire (Omar, 2018) mais on peut tenter de la formuler comme étant un effort multidisciplinaire d'intégration et d'optimisation des structures, systèmes et services d'un bâtiment dans le but de répondre aux enjeux généraux de confort, de bien-être, de sécurité (pour les usagers), de rentabilité, de performance énergétique et environnementale (pour l'exploitant), en s'appuyant sur des technologies liées au monde numérique (Arkin, Paciuk, 1997)(Architecture et Climat, 2019).

La composante SI y occupe un rôle crucial puisqu'elle doit assurer cette « intelligence » en permettant aux différents objets, capteurs et fonctions du bâtiment d'interagir avec la possibilité de contrôle à distance. Une évolution importante par rapport à une pure Gestion Technique Centralisée (GTC) est le contexte ouvert et décentralisé (Architecture et Climat, 2019). On a affaire à un ensemble de SI en charge de divers sous-systèmes qui doivent collaborer de manière flexible à la fois en interne (avec les utilisateurs, les gestionnaires, la maintenance) mais aussi avec de nombreux autres systèmes externes (fournisseurs de services énergétiques/ eau, communication, intervention d'urgence...) comme illustré à la figure 1.

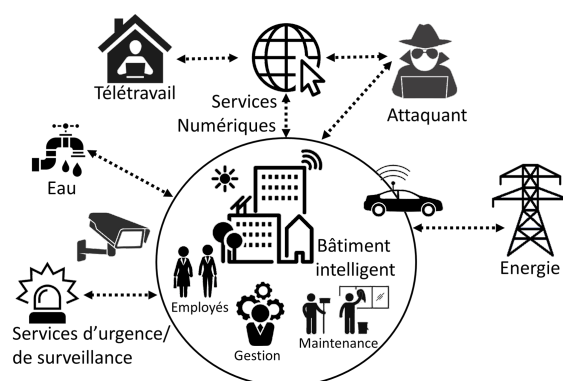


Figure 1. Principaux intervenants dans un bâtiment intelligent

Ces SI sont pervasifs et prolifèrent pour aller s'ancrer dans divers dispositifs tels que les systèmes électriques (compteurs, prises, appareils connectés), les systèmes d'alarme, le contrôle d'accès, les véhicules des usagers, les usagers eux-mêmes (p.ex.

via leurs smartphones) en utilisant des systèmes relais qui gagnent en capacité de traitement (border router « Edge »). Certains services peuvent aussi être externalisés dans le nuage (cloud). Tout ceci expose aussi ces SI à des attaques comme représenté sur la figure 1.

Cet article examine les SI présents dans les bâtiments intelligents. Il s'intéresse à tout le cycle de vie d'un bâtiment qui inclut sa conception, sa construction, son opération, voire son démantèlement. L'objectif est de conduire une analyse SWOT (Force-Faiblesse-Opportunités-Menaces) liée à l'évolution vers des systèmes de plus en plus pervasifs. A cette fin, nous structurons notre démarche comme suit : Dans la section 2, nous détaillons d'abord les différentes classes de besoins généralistes, évoquées dans la définition puis identifions différents types de SI présents dans ce domaine. La Section 3 présente des éléments d'une étude de cas réalisée dans un centre de recherche hébergeant un des auteurs. Dans la section 4, nous présentons notre analyse SWOT qui s'appuie sur le matériel des sections précédentes et des travaux de la littérature. Enfin, la Section 5 tire quelques conclusions et pistes d'élaboration de notre travail qui ne prétend aucunement être exhaustif à ce stade mais vise surtout à fournir un cadre intéressant pour alimenter l'atelier.

2. Caractérisation des SI pour bâtiments intelligents

2.1. Catégories d'exigences

Bien que chaque bâtiment ait une conception relativement spécifique liée à son affectation, ces bâtiments partagent tous un ensemble de capacités de base :

- connecter les humains dans des espaces physiques mais également virtuels
- leur permettre de collaborer efficacement pour mener à bien leurs missions
- assurer un meilleur contrôle des installations et des opérations
- gérer de manière durable des ressources telles que l'espace, l'énergie, l'eau et les employés.

Nous développons rapidement une série d'exigences non-fonctionnelles (NFR) qui impactent largement la conception et le fonctionnement de ces systèmes et leur SI. Celles-ci font partie des exigences types de systèmes pervasifs (Khalfi, Benslimane, 2014) mais aussi plus spécifiques au domaine considéré (Kleissl, Agarwal, 2010)(Lawrence *et al.*, 2016)(EC, 2017)(Kaufmann *et al.*, 2018).

NFR de rentabilité : maintenance et gestion énergétique

La rentabilité d'un bâtiment est un point clef. On estime que seulement 20% du coût total de possession (TCO) est lié à sa construction et les 80% représentent des coûts d'opérations (dont 40% sont des coûts énergétiques et 30% des coûts de maintenance) (EC, 2017).

L'intérêt de la maîtrise de l'efficacité énergétique est à la fois économique et environnemental puisqu'il permet de réduire les émissions de CO2 liées à l'utilisation

des énergies fossiles. Cependant, certaines actions relèvent de la phase de conception (isolation par exemple) tandis que d'autres peuvent se baser sur une optimisation des ressources, par exemple, en fonction de la présence (lumière, air conditionné, etc.).

La maintenance peut être optimisée sur base d'une connaissance de l'état de différents sous-systèmes, des besoins préventifs d'entretien et d'indicateurs de dégradation.

NFR de sécurité physique et de confort des utilisateurs

La sécurité physique et plus largement le confort sont des aspects très importants pour les utilisateurs afin qu'ils soient dans un environnement non seulement sûr mais aussi confortable et propice à la collaboration (Abdennadher *et al.*, 2015).

Afin que l'environnement soit sûr, les principales exigences sont relatives aux dispositifs d'évacuation d'urgence et de contrôle d'accès.

A des fins de confort, on veille à fournir des conditions environnementales optimales en termes de température, humidité, nuisances sonores, etc.

NFR Ouverture et flexibilité

L'aspect ouvert est une caractéristique importante d'un bâtiment intelligent qui entre en interaction avec un environnement externe dynamique. Ainsi, au niveau énergétique, il doit être capable de s'ajuster aux conditions du réseau en privilégiant les heures où l'énergie est abondante voire produite localement (ex. panneaux photovoltaïques). Cela nécessite de pouvoir planifier des actions, par exemple, pour enclencher ou délester certains dispositifs électriques liés à une forte inertie thermique.

La flexibilité concerne l'adaptation à de nouvelles exigences suite à une évolution du métier ou de l'environnement. La crise de la COVID-19 illustre des changements radicaux avec un basculement vers un mode de travail réduit dans les locaux et une augmentation importante du recours aux infrastructures IT pour télétravailler. Cela impacte aussi d'autres systèmes tels que le chauffage ou la planification des salles de réunion afin de tenir compte des capacités maximales revues à la baisse.

2.2. Cycle de vie

Le cycle de vie d'un bâtiment intelligent peut se découper en trois grandes phases : la conception, la construction et l'opération. Chaque phase dispose d'un propre référentiel de modélisation dans laquelle s'inscrit son SI (BIMTonic, 2021). La figure 2 illustre les importances relatives en lien avec les coûts évoqués précédemment.

– la phase BIM (Building Information Modeling, en français : Bâti Immobilier Modélisé) concerne la phase de conception par l'architecte. Elle consiste en l'examen et la modélisation 3D des différentes facettes ainsi qu'une première estimation des coûts. Le SI associé concerne la modélisation et la simulation du bâtiment en devenir ou à rénover.

– la phase BAM (Building Assembly Modeling) débute à l'adjudication du marché. L'entrepreneur enrichit le modèle avec des informations pratiques liées à l'exé-

cution. Le SI impliqué ici assure le suivi de chantier et permet d'organiser au mieux un planning, la logistique, le stockage et ainsi, réduire les coûts de construction.

– la phase BOOM (Building Operation Optimisation Model) débute à la livraison. Elle est assurée par le maître d'ouvrage. Elle vise à améliorer et optimiser l'utilisation du bâtiment notamment en termes de consommation énergétique et de maintenance. Les SI impliqués forment à la fois le système nerveux du bâtiment, collectant les informations et contrôlant ses fonctions, mais aussi son cerveau pour en coordonner le fonctionnement. Ces SI sont pervasifs au sens physique mais via leur articulation avec les SI du BAM et BIM, il y a aussi une continuité temporelle.

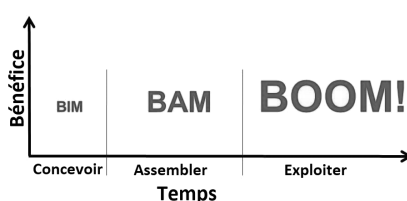


Figure 2. Importances relatives des phases du cycle de vie

3. Étude de cas

Cette section décrit deux scénarios de déploiement de SI dans notre bâtiment mené par le projet Build4Wal, en collaboration avec le domaine de la construction en Belgique (Bernaud, 2019)(CSTC, 2019). Cette section plus technique permet d'avoir une vue complémentaire aux objectifs généraux présentés à la Section 2.1.

3.1. Scénario de données de confort

L'objectif de ce scénario est de déployer un réseau de capteurs afin de collecter des informations sur l'environnement du bâtiment en les plaçant dans les différentes pièces à travers les 3 étages tels que représentés par la figure 3. Le bâtiment présente de nombreuses contraintes comme des murs en béton armé ou encore des installations électriques pouvant créer des interférences. Les données récoltées concernent la température, la luminosité, l'humidité ou encore la pression atmosphérique. Les critères clés de sélection des capteurs IoT concernent leur consommation et la portée.

Le choix de conception s'est porté ici sur une architecture basée sur Z-Wave satisfaisant les deux critères malgré un coût relativement élevé des capteurs et une couche applicative peu adaptable. À contrario, le protocole 6LoWPAN possède des performances similaires, et permet également d'adapter les applications grâce à des requêtes CoAP/IPv6-compressé. Cette adaptation permet à n'importe quel capteur compatible avec le protocole 802.15.4 de communiquer avec une plateforme cloud. Il supporte aussi une topologie « mesh » permettant de couvrir tout le bâtiment avec une architecture qui se (re)configure dynamiquement. Un traitement spécifique est réalisé en bordure du réseau de capteurs sur le routeur 6LBR qui peut embarquer des fonctions

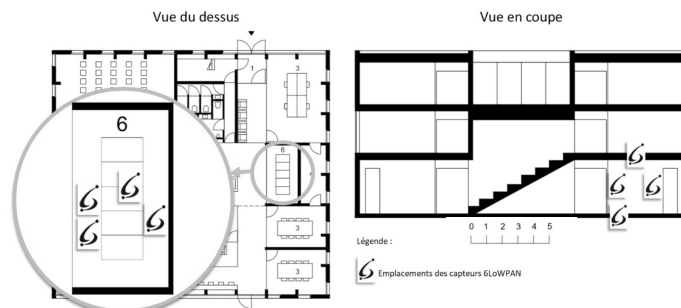


Figure 3. Scénario de collecte de données de confort

de pré-traitement sur ce « Edge » avant la remontée des données vers la plateforme de traitement. Cette dernière est constituée d'un collecteur de données (Prometheus) couplé à un outil de dashboard (Grafana). Des traitements plus poussés sont aussi possibles via une plateforme d'analyse prédictive permettant d'optimiser le pilotage d'équipements tels que la pompe à chaleur du bâtiment.

3.2. Scénario de gestion énergétique

Dans ce scénario, le SI doit assurer le monitoring des « gros postes » de consommation : HVAC, l'armoire électrique principale et le bloc d'alimentation qui contient les compteurs officiels d'eau, de gaz et d'électricité. Le réseau s'étend sur tout le bâtiment et sur le parking, soit une surface totale de 3800 m². Deux technologies sont candidates pour le monitoring des « gros postes » de consommation : WM-Bus et LoRa. Elles possèdent les mêmes caractéristiques en termes de portée, de consommation et de débit. Les deux différences majeures sont le prix et l'évolutivité. WM-Bus est un protocole très structuré, ce qui facilite le déploiement, mais le prix d'acquisition de capteurs reste très élevé. La technologie LoRa est plus abordable et plus facile à adapter à un format de données. Il est même possible de reproduire la structure de WM-Bus et c'est pour ces raisons qu'il a été préféré. Le réseau de capteurs est basé sur une topologie arbre-étoile. Les capteurs envoient leurs données en LoRa à une passerelle via le protocole MQTT qui permet aux SI de s'abonner à des messages.

4. Analyse SWOT

L'évolution vers des bâtiments intelligents s'appuie sur des SI de plus en plus pervasifs. Elle présente des bénéfices mais génère également des risques et de nouveaux défis. **Afin de les identifier, cette section présente une analyse SWOT qui a été conduite selon le processus suivant** : A partir des diverses exigences présentées à la section 2.1, nous avons distingué les forces/faiblesses (internes au bâtiment sur base du contexte établi à la figure 1) ainsi que des opportunités/menaces (externes au bâtiment) sur base du contexte établi à la figure 1. Notre analyse a été alimentée par notre étude de cas et par des travaux rapportés dans la littérature.

4.1. Forces (Strengths)

Meilleure coordination de bout en bout d'un projet immobilier. Grâce à l'enchaînement et l'enrichissement des modèles en phase de conception, construction et opération, l'information est transmise plus efficacement (Drăgoicea *et al.*, 2013). Il en résulte une meilleure estimation des matériaux nécessaires et planification des travaux. La sécurité sur le chantier est aussi mieux préparée (Kaufmann *et al.*, 2018).

Gestion optimisée des coûts d'opération et contrôle de l'empreinte écologique. Le contrôle des coûts d'opération résulte à la fois d'une optimisation des systèmes, des services et de la gestion (Omar, 2018). Cela intervient au niveau de la phase de conception mais aussi dynamiquement lors du fonctionnement afin de ne pas gaspiller d'énergie, notamment en tenant compte de l'activité des utilisateurs (Nguyen, Aiello, 2013). La finesse de la collecte des informations via les capteurs de type IoT permet de développer une gestion fortement contextualisée (Elkhokhi *et al.*, 2018). Globalement, ceci contribue à la réduction de l'empreinte écologique.

Facilité de configuration et d'adaptation. Le recours à une couche intelligente pilotée par logiciel augmente les possibilités de coordination avec une grande flexibilité et adaptabilité à l'évolution des usages, y compris dans des scénarios imprévus (par exemple, la crise COVID-19). La dimension pervasive nécessite cependant une attention particulière à l'architecture déployée notamment pour prendre en compte l'hétérogénéité, le contexte et la dynamique (Khalfi, Benslimane, 2014).

4.2. Faiblesses (Weaknesses)

Mauvaise expérience utilisateur par rapport aux systèmes mis en place. Les obstacles peuvent être de nature différente : système non fiable (ex. contrôle d'accès), impression d'être espionné (ex. gestion des informations de présence), mauvaise interface utilisateur de services, mauvaise organisation du bâtiment lui-même, etc. Ce type de faiblesse peut-être adressé en impliquant les utilisateurs plus tôt dans le processus, dès la conception (BIM), en validant des affinements (dans le BAM) et pour les aspects de personnalisation au niveau du BOOM.

Perte de maîtrise causée par une application trop poussée de l'automatisation, l'optimisation et la prédiction combinées avec une réduction des possibilités de s'écarter des décisions prises par les SI (Architecture et Climat, 2019). On peut se retrouver dans des situations où le bâtiment décide de baisser les volets automatiques pour allumer les lumières alors que les utilisateurs désirent profiter du soleil. Inversement, des limites d'automatisation de processus de configuration peuvent engendrer des opérations fastidieuses et coûteuses, par exemple, en cas de mise à jour nécessaire ou de remplacement de certains capteurs. Les protocoles adaptatifs en « mesh » évoqués dans l'étude de cas permettent aussi de pallier ce type de problème.

Pérennité des informations. Certaines informations doivent être conservées sur toute la durée de vie du bâtiment de plusieurs décennies, notamment les modèles BIM/BAM/BOOM dont les informations restent utiles au démantèlement. La pré-

servation de données informatiques à long terme nécessite une approche réfléchie (Aliouali, Dahmane, 2007) et basée sur des standards tels que OAIS (ISO, 2012).

Amortissement du surcoût. Le bâtiment intelligent permet de rationaliser la consommation d'énergie et d'atteindre une meilleure efficacité environnementale mais au prix d'un surcoût technologique dans les phases de conception, de construction et d'opération. Il convient donc de veiller à ce que ce surcoût soit amorti sur la durée de vie du bâtiment et ne soit pas grevé par des coûts imprévus, par exemple, liés à des composants plus énergivores qu'anticipé ou à de l'obsolescence affectant le SI (voir menaces).

4.3. *Opportunités (Opportunities)*

L'ouverture des SI sur les systèmes externes permet d'envisager des évolutions vers des scénarios d'intégration plus riches entre le bâtiment et des systèmes externes. Par exemple, en cas de besoin d'intervention d'urgence, des informations précises collectées peuvent être transmises en temps réel à des intervenants autorisés et des permissions d'accès peuvent être activées. Au niveau énergétique, le bâtiment peut devenir un composant participant activement dans un smart grid en permettant notamment la prise de contrôle sur certains appareils afin de les activer au moment le plus opportun (Maasoumy *et al.*, 2015)(Lawrence *et al.*, 2016).

Utilisation de techniques d'IA. Les techniques récentes d'intelligence artificielle peuvent également être envisagées afin de permettre une optimisation plus poussée des systèmes en utilisant des techniques d'apprentissage, permettant au bâtiment de mieux anticiper les patterns d'utilisation des locaux et services. Des techniques d'apprentissages permettent de prédire assez précisément les chances d'occupation en fonction de plusieurs paramètres de temps et de durée (Yu, 2010). Des modules de simulation utilisés au niveau du BIM peuvent aussi évoluer en véritable jumeau numérique utilisable en phase d'exploitation afin d'aider à la prise de décision (Cohen Boulakia *et al.*, 2020). Ceux-ci peuvent aussi être couplés à des données de systèmes externes, par exemple, les prévisions météorologiques (Kavalionak, Carlini, 2019).

4.4. *Menaces (Threats)*

Cybersécurité. L'augmentation de l'interconnexion et l'infiltration des SI dans toutes les fonctions du bâtiment intelligent accroît fortement la surface d'attaque. Il est impératif que des mesures de cybersécurité soient mises en place pour traiter ces risques, notamment via une analyse de risques dès le stade de la conception. Le modèle BIM peut servir de support pour découvrir des scénarios de menaces et évaluer les impacts. La nature du déploiement des SI doit être intégrée à cette analyse en traçant le parcours de l'information depuis des sources et les différents traitements qui peuvent inclure des composants proches des capteurs (Edge) jusqu'à des parties externalisées dans le Cloud. Ces deux menaces sont présentes dans notre étude de cas et également discutées dans (Pan, McElhannon, 2018). Les mesures peuvent ensuite être intégrées

dans la mise en oeuvre du BAM et le suivi opérationnel BOOM où ils doivent faire l'objet d'une surveillance. Un point d'attention spécifique est la protection des données personnelles collectées de diverses façons (contrôle d'accès, capteur de présence, authentification sur des postes...) en respect du RGPD (EC, 2016). On peut augmenter le modèle BOOM avec la notion de propriété « ownership » des données et fournir des garanties aux utilisateurs sur leur utilisation (Boyer *et al.*, 2006).

Dépendance technologique. Note étude de cas illustre la difficulté de mise en oeuvre du BAM. Les critères considérés (coût, consommation, adaptation...) peuvent conduire à différents choix. Les standards ouverts sont à privilégier, de même que les technologies plus matures et sécurisées mais face à des technologies en évolution rapide, des compromis sont généralement nécessaires. Privilégier des solutions aisément adaptables permet aussi de gérer cette évolution en maîtrisant les coûts.

5. Conclusion et Perspectives

Cet article propose une analyse SWOT de la transformation digitale menant à l'émergence de bâtiments intelligents qui puisent leurs informations dans un réseau de capteurs et d'interfaces connectées à de multiples SI, permettant ainsi, d'en gérer et optimiser le fonctionnement. Pour mener notre analyse, nous avons identifié les principaux enjeux, les intervenants ainsi que le type de SI impliqués. Nous avons aussi puisé dans un cas d'étude et dans des retours d'expérience de la littérature. Nous avons tenté d'identifier des pistes actuellement identifiées pour exploiter au mieux le potentiel d'innovation et limiter l'impact de menaces.

L'analyse réalisée ne prétend pas être exhaustive mais nous espérons qu'elle donnera une esquisse assez représentative des enjeux et défis relatifs aux SI pour alimenter les réflexions de la communauté. Par la suite, nous comptons intégrer plus de retours en élargissant notre étude de cas (trop centrée sur la phase d'opération et sur l'infrastructure de capteurs), en intégrant une revue plus poussée de la littérature et des retours collectés lors de plusieurs ateliers. Nous comptons aussi examiner l'évolution vers des technologies de systèmes autonomes à la fois du point de vue du SI et du point de vue énergétique.

Bibliographie

- Abdennadher I. *et al.* (2015). *Designing energy efficient smart buildings in ubiquitous environments*. In 15th Int. Conf. on Intelligent Systems Design and Applications (ISDA).
- Aliouali N., Dahmane M. (2007). *La préservation des documents numériques : enjeux et stratégies*. RISTVol, vol. 17, n° 1-2.
- Architecture et Climat*. (2019). Les dossiers thématiques : Smartbuilding.
- Arkin H., Paciuk M. (1997). *Evaluating intelligent buildings according to level of service systems integration*. Automation in Construction, vol. 6, n° 5, p. 471-479.
- Bernaud B. (2019). Build4Wal - Construction 4.0. *Rapport scientifique CETIC*.

- BIMTonic.* (2021). Qu'est-ce que le 'bim bam boom'. <https://www.bimtonic.be/fr/expertise-bim/faq/35/quest-ce-que-le-bim-bam-boom>.
- Boyer J. P., Tan K., Gunter C. A.* (2006). *Privacy sensitive location information systems in smart buildings*. In *Security in pervasive computing*. Springer Berlin Heidelberg.
- Cohen Boulakia B., Finance B., Chevallier Z.* (2020, 05). *A reference architecture for smart building digital twin*. In *Sedit, heraklion, greece*.
- CSTC.* (2019). BUILD4WAL Centre de Démonstration Construction numérique. <https://www.cstc.be/homepage/index.cfm?cat=projects&proj=1368>.
- Drăgoicea M., Bucur L., Pătrașcu M.* (2013). *A service oriented simulation architecture for intelligent building management*. In *Exploring services science*, p. 14–28.
- EC.* (2016). Regulation (EU) 2016/679 - General Data Protection Regulation (GDPR). <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679>.
- EC.* (2017). *Smart Building: Energy efficiency application*. <https://ati.ec.europa.eu/reports/sectoral-watch/smart-building-energy-efficiency-application>.
- Elkhoukhi H. et al.* (2018). *Towards a real-time occupancy detection approach for smart buildings*. *Procedia Computer Science (MobiSPC 2018)*, vol. 134, p. 114-120.
- ISO.* (2012). *14721 - Reference Model For An Open Archival Information System (OAIS)*. <https://www.iso.org/standard/57284.html>.
- Kaufmann D., Ruaux X., Jacob M.* (2018). *La digitalisation du secteur de la construction: la révolution est en marche*. Cabinet Oliver Wyman.
- Kavalionak H., Carlini E.* (2019). *An hvac regulation architecture for smart building based on weather forecast*. In *Economics of grids, clouds, systems, and services*. Springer Int. Pub.
- Khalfi M., Benslimane S. M.* (2014, 06). *Systèmes d'information pervasifs : Architecture et challenges*. In *10èmes journées francophones mobilité et ubiquité*.
- Kleissl J., Agarwal Y.* (2010). *Cyber-physical energy systems: Focus on smart buildings*. In, p. 749–754. Association for Computing Machinery.
- Lawrence T. M. et al.* (2016). *Ten questions concerning integrating smart buildings into the smart grid*. *Building and Environment*, vol. 108, p. 273-283.
- Maasoumy M., Nuzzo P., Sangiovanni-Vincentelli A.* (2015). *Smart buildings in the smart grid: Contract-based design of an integrated energy management system*. In *Cyber physical systems approach to smart electric power grid*, p. 103–132.
- Nguyen T. A., Aiello M.* (2013). *Energy intelligent buildings based on user activity: A survey*. *Energy and Buildings*, vol. 56, p. 244-257.
- Omar O.* (2018). *Intelligent building, definitions, factors and evaluation criteria of selection*. *Alexandria Engineering Journal*, vol. 57, n° 4, p. 2903-2910.
- Pan J., McElhannon J.* (2018). *Future edge cloud and edge computing for internet of things applications*. *IEEE Internet of Things Journal*, vol. 5, n° 1, p. 439-449.
- Yu T.* (2010). *Modeling occupancy behavior for energy efficiency and occupants comfort management in intelligent buildings*. In *9th int. conf. on machine learning and applications*.

Modeling Energy Consumption in SOA: Requirements and Current Status

**Jorge Andrés Larracochea¹, Philippe Roose², Sergio Ilarri³,
Yudith Coromoto Cardinale⁴, Sébastien Laboire⁵, Mauricio
Jacobó González⁶**

1. LIUPPA/E2S, Université de Pau et des Pays de L'Adour
2 Allée du Parc de Montauray, 64100 Anglet, France
jorge-andres.larracochea@etud.univ-pau.fr
2. LIUPPA/E2S, Université de Pau et des Pays de L'Adour
2 Allée du Parc de Montauray, 64100 Anglet, France
Philippe.Roose@iutbayonne.univ-pau.fr
3. Instituto de Investigación en Ingeniería de Aragón /I3A, Universidad de Zaragoza
Mariano Esquillor s/n, 50018, Zaragoza, España
silarri@unizar.es
4. Dpto. de Computación y T.I, Universidad Simón Bolívar
1080-A Caracas, Venezuela
ycardinale@usb.ve
5. LIUPPA/E2S, Université de Pau et des Pays de L'Adour
2 Allée du Parc de Montauray, 64100 Anglet, France
Sebastien.Laborie@iutbayonne.univ-pau.fr
6. LIUPPA/E2S, Université de Pau et des Pays de L'Adour
2 Allée du Parc de Montauray, 64100 Anglet, France
mj.gonzalez@etud.univ-pau.fr

ABSTRACT. Even though hardware architects have managed to incrementally mitigate energy consumption in information and communication technology devices, energy consumption will always be a requisite for software execution. This has motivated researchers to develop a limited amount of methodologies that promote green software development and its philosophy, with new assessment methods for calculating the energetic costs of software development and software execution. In spite of this, they have been acknowledged and adopted with limited success, as they try to address highly-volatile variables (like human behavior) and environments with specific hardware/software platforms and language-centric solutions. This has created a conflict between theory and practice where, otherwise, a generic and adaptive approach could manage the discord, especially in Service-Oriented

Architectures (SOA). In this paper, we present the works available in relation to services' requirements definition and profiling for energy management, as well as their limitations and advantages in relation to green software development. Furthermore, we identify the key limitations in existing proposals and the challenges that need to be tackled to enable an effective and developer-friendly energy-aware development of SOA-based applications.

KEYWORDS: Energy management, Service-Oriented Architecture, Green IT, Energy Consumption, Services.

1. Introduction

Nowadays, an enormous amount of ICT (Information and Communication Technologies) devices dedicated to both consumer and enterprise markets are available in the network. For instance, the number of smartphones sold to end users in 2020 alone was 1,540.00 millions of units (Statista, 2020). Even though the energy consumption of hardware components has been drastically mitigated in the last decade, such as the reduction of electrical consumption per gigabyte of mobile data transmitted from 12.34 kWh in 2010 to an estimated magnitude < 0.1 kWh in 2020 in Finland alone (Pihkola et al., 2018), software will always rule the energy consumption in hardware, as hardware's purpose is to process software's demands. Due to a high level of heterogeneity in both hardware and software platforms, and an increasing research interest on the problem of energy consumption, ICT researchers are concerned with controlling the behavior of software and how it is conceived, segmenting these topics into a branch of studies called Green Software Development (GSD). Some of these efforts have produced energy-saving software development methodologies and methods that can be divided in three main categories (Acar, 2017): green with software, green within software, and green software. *Green with software* sets the goal as the creation of software that provides frugal solutions to external problems; these types of solutions consider variables surrounding the problem and their energetic impacts in order to generate the "greenest" (most frugal) solution possible. *Green within software*, on the other hand, seeks to reduce the power consumption of software using a power model. *Green software* development methodologies establish guidelines for all the variables surrounding and included in the Software Development Life Cycle (SDLC), such as the source of electricity or energy-efficient development equipment, to result in a "greener" software development that reduces the negative impacts on society, economy, and the environment. Several approaches have contributed in this field with solutions for assessing the energy consumption of software from the development to the deployment/maintenance phases of the SDLC. Some of them focus on trying to understand the impact of system settings (Peltonen et al., 2018), assessing or ranking the consumption of apps (Behrouz, 2015 ; Oliner, 2013 ; Pathak, 2012), analyzing power consumption with testbeds (Hindle et al., 2014), performing code analysis (Aggarwal, 2015 ; Hao, 2013 ; Manotas, 2014 ; Pathak, 2012), and analyzing the consumption of system calls in the evolution of software (Aggarwal et al., 2015). They are positive contributions, as they provide proper methods to understand how energy is consumed, and without them hardly any

measure could be applied to assess and manage it. However, they have several caveats, as they rely on extra hardware, specific techniques, and specific platforms. Overall, their analysis is usually performed during or after the development phase; extra time invested in energy efficiency has been frowned upon by stakeholders before (Jagroep et al., 2017), even though a survey concluded that 65% of experienced people in the field of software development identified the efficient utilization of time and computing resources as a success factor within agile software development (Rashid et Khan, 2017). As seen in Fig. 1, the research from the development to the deployment/maintenance phases of the SDLC has been mostly taken care of regarding energy assessment and management, but there is a demonstrated lack of research work that provides guidelines for the analysis phase (Georgiou et al., 2019). Furthermore, approaches concerning the design phase still need more work, as some concluded with negative energy savings in embedded and Android-based systems, due to erroneous preconceptions of design patterns that were thought efficient (Bunse, 2013 ; Sahin, 2012).

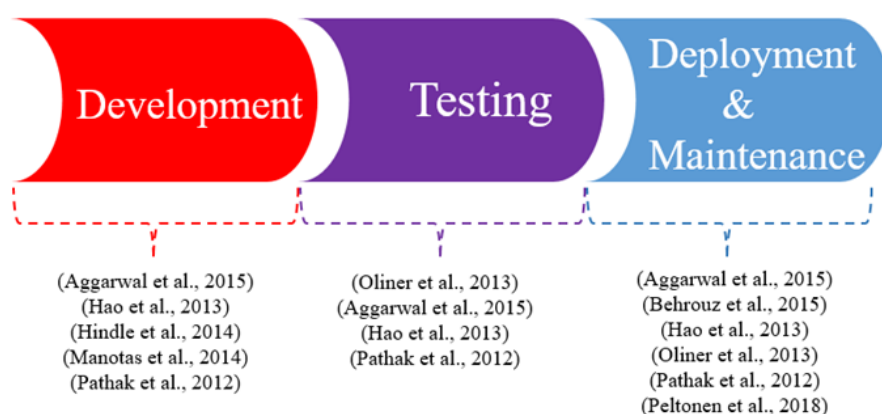


Figure 1. Distribution of works available along the SDLC

This is why in this article we analyze the state of the art in this field and identify key limitations that need to be addressed to advance work towards better energy management by closing the scientific gap among analysis, design, and the rest of the SDLC. We mainly focus on Service Oriented Architectures (SOA), due to their flexibility, current widespread use, and seemingly rising popularity. The goal of this work is to provide a study of the advantages and shortcomings of existing literature on SOA, and analyse their research in regards to service-oriented requirements engineering, evaluation of energy consumption, green software, and energy management.

In Section 2, we present background concepts on applications and SOA. In Section 3 we describe previous works regarding energy management for SOA and

our findings, we present our future work in section 4 and, finally, our conclusions in section 5.

2. Background

According to the Oxford dictionary, an application is a piece of software that fulfills a specific purpose. An application can be conceived with any architectural pattern that suits its context, and it is defined during the design phase of the SDLC. The purpose of architectural patterns is usually to “address an application-specific problem within a specific context and under a specific set of limitations” (Pressman et Maxim, 2014). For the purpose of our work, we will focus on the Service Oriented Architecture (SOA); a SOA consists of an application composed of smaller units of software called services. Each service consists of the code and data to perform a specific business function and their interfaces provide a loose coupling (IBM Cloud Education, 2019). This architecture provides advantages such as reuse of pre-existing services, easy discoverability of services offered to consumers and abstraction from the physical implementation of a service due to interfaces. Services, however, offer a wide range of other properties that could be described in relation to characteristics they exhibit, such as the ones shown in Fig. 2.

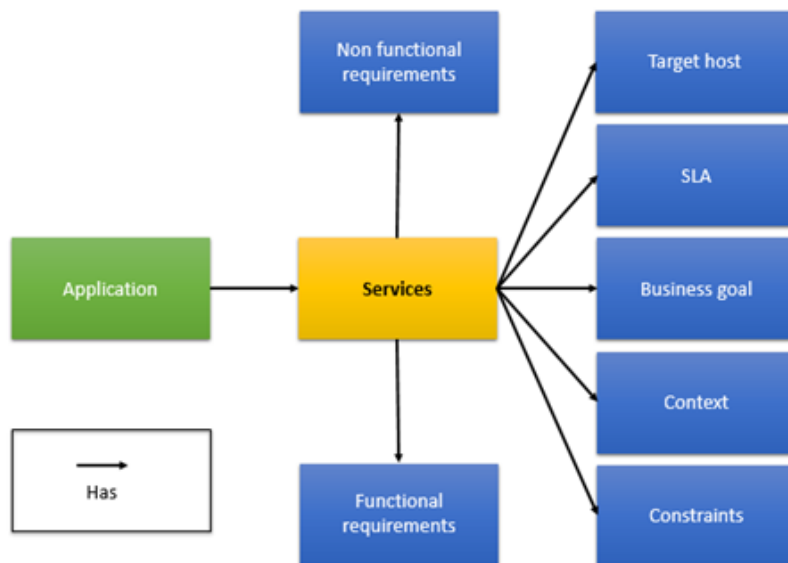


Figure 2. Other services' properties

These characteristics allow software designers/developers to fully abstract all the properties that describe a service in order to profile it for its identification and discovery. They can also be seen from three different perspectives in relation to the

main roles available in SOA: service provider, service requester (or consumer), and service registry. The service provider publishes services to the service registry and the service registry works as an intermediary between the service publisher and the requester, aiding the service requester in the discovery and identification of published services.

3. Previous Works

3.1 Software Requirements Engineering in SOA

SOA is by no means a new approach. Some of the biggest IT companies such as IBM and Oracle have backed this architecture by creating frameworks for it, such as the Oracle Service Engineering Framework (Oracle, 2012) pushing its adoption within the industry. Furthermore, ISO/IEC 18384-2:2016 (ISO, 2016) provides a fully standardized reference architecture for design, performance, development, deployment, and management of SOA solutions. It also provides useful categories of services to better classify them by their purpose. Even though both the framework and the ISO/IEC reference architecture provide useful and thorough guidelines, they do not provide a specific set of functional and non-functional requirements that depict neither a qualitative or quantitative consumption of resources that services exhibit. This is also applicable to other approaches that suggest quality attributes to evaluate services and SOA (Bianco, 2007 ; O'Brien, 2005).

3.2 SOA Quality Attributes and Consumption Evaluation

Other approaches have conceived further non-functional properties (NFP) to evaluate services from the consumer perspective, such as the ones suggested by Becha et al., where a mention of possible units of measurement in relation to these properties can be found (Becha et Amyot, 2012). That work is a step forward towards an actual evaluation of services' behavior and an aid in service selection. The study, however, concludes that a formalization of how each NFP can be described and measured is needed. Choi et al. contributed by modeling the QoS attributes and metrics for services evaluation in SOA from the perspective of the service consumer (Choi et al., 2007). As we can see, contributions have been made around quality assurance of services and Non-Functional Properties (NFP) definitions without talking about actual resource management, energy consumption model, or profiling of services.

3.3 Green Software and Energy Management

The topic of "green software" or energy management is, overall, not mentioned in the literature related to SOA, as the review conducted by Niknejad et al. (Niknejad et al., 2020) demonstrates. A notable exception absent in this review is the work published in 2015 by Ibrahim Naseem, ranking energy-aware services

(Ibrahim, 2015). He, in contrast to other works, introduces an Energy-Aware Service Oriented Architecture (EASOA) that provides formal definitions for Energy-Aware Services (EAS) using a model-based specific notation. An Energy-Aware Service Mapper takes the requests and provides matches that meet the requester's requirements. He also introduces a ranking algorithm that considers the amount of energy consumed by a service in the ranking process. As we can see, Ibrahim Naseem provides definitions and a matching approach according to energy consumption and requirements, which we find interesting. However, the definition process requires deep knowledge from the designer and the approach has not been validated.

3.4 Data Management and Energy Consumption in SOA

Throughout the works provided and described in previous sections, scarce mentions are made about profiling of data and the energetic cost operations with/on it can incur (Oracle, 2012). Even though a comment about increased energy consumption when the distance between the service provider and consumer is significant can be found (Ibrahim, 2015), no further mention about techniques, profiles or models of data's consumption of resources are profoundly discussed in any of them. Existing literature that provides properties for data profiling or analysis is already available (Patgiri, 2016 ; Wang, 2011), but a link between SOA and energy management/assessment/profiling of data is still missing.

4. Future Work and Associated Challenges

After assessing the advantages and shortcomings of available literature, it is our ambition to close the scientific gap found. We aspire to achieve this by continuing our work with the future objectives and their associated challenges described below.

4.1 Define a set of properties to create a resources consumption profile of services according to their requirements

In order to properly understand the consumption of resources and behavior a service presents, we need to profile it according to its requirements. This will not only aid in the analysis phase of the SDLC and Service Oriented Requirement Engineering (SORE), it will provide a wider vision to the designer/developer of the role the profiled service plays in the greater scheme of the application and whole architecture. It poses a major challenge as properties required to create a profile must be general enough to be applicable for any type of service.

4.2. Define a set of properties to create a resources consumption profile of data

In the same fashion as services, a profile of data will aid the designer/developer to understand how it flows among services and applications, and provide us with

knowledge of how it is used/generated. This is one of our main objectives due to the lack of research performed in this area, especially in relation to their impact in energy management of SOA applications.

4.3. Create an intuitive workflow for generating profiles

In order to avoid a steep learning curve or difficult introduction of designers/developers to the creation of profiles, we aim at creating an intuitive and easy to use workflow for data and services profile creation, especially targeted at non-expert designers and developers. This will aid us to gain traction in the adoption of our approach, making it accessible to anyone interested.

4.4. Create models of resources consumption based on profiles

Even though services and data profiles might provide us with useful descriptions of the consumption of resources they present (among other characteristics), it is not enough to perform proper energy management. As it is the cornerstone of our approach, we will assess the profiles of data and services to generate models that can be further evaluated and used towards energy management.

4.5. Assess models and in order to categorize their consumption

An important aspect of models is being able to categorize the results they provide. We aim at creating a descriptive “identity tag” based on the assessment of models and standardizing them in order to facilitate the identification and discovery phase of SOA with a focus in energy management. Furthermore, we aspire to generate these identity tags at any level of evaluation, from services to whole applications, as well as provide an identity tag that developers and consumers of applications can understand.

4.6. Discover strategies to manage the consumption of models

As models will provide us with actual information on consumption of resources and profiles other characteristics such as dependence among services and data, we will be able to create strategies fit to both. The goal of these strategies is performing energy-aware management of services and data. These strategies are a challenge in itself due to the validation process and constraints possible within the models and profiles.

5. Conclusions

We have found works on SOA standardization, SOA applications’ quality evaluation, and service-oriented requirements engineering are available, but an

emphasis in energy management or green software as a goal in SOA is fairly absent. Furthermore, the lack of research related to the energy/resources consumption caused by data consumption/generation in SOA applications increments the gap in this field, and presents us with an opportunity for new approaches to fulfill the current needs of better management of energy consumption, especially from the analysis and design phase of the SDLC. We have also presented our future work and challenges to contribute to the research in this topic with the hope of filling in the gap and aid researchers, designers and developers to study/design/develop more energy-aware SOA applications.

References

- Aggarwal, K., Hindle, A., Stroulia, E.: *GreenAdvisor: A tool for analyzing the impact of software evolution on energy consumption*. In: 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME). pp. 311–320. (2015)
- Becha, H., Amyot, D.: *Non-Functional Properties in Service Oriented Architecture – A Consumer’s Perspective*. J. Softw. 7 (2012)
- Behrouz, R.J., Sadeghi, A., Garcia, J., Malek, S., Ammann, P.: *EcoDroid: An Approach for Energy-Based Ranking of Android Apps*. In: 2015 IEEE/ACM 4th International Workshop on Green and Sustainable Software. pp. 8–14. (2015)
- Bianco, P., Kotermanski, R., Merson, P.F.: *Evaluating a Service-Oriented Architecture*. (2007)
- Bunse, C., Gottschalk, M., Naumann, S., Winter, A.: 2nd Workshop EASED@BUIS 2013 - *Energy Aware Software-Engineering and Development*. Presented at the Softwaretechnik-Trends, April 25 (2013)
- Choi, S.W., Her, J.S., Kim, S.D.: *Modeling QoS Attributes and Metrics for Evaluating Services in SOA Considering Consumers’ Perspective as the First Class Requirement*. In: The 2nd IEEE Asia-Pacific Service Computing Conference (APSCC 2007). pp. 398–405. (2007)
- Georgiou, S., Rizou, S., Spinellis, D.: *Software Development Lifecycle for Energy Efficiency: Techniques and Tools*. ACM Comput. Surv. 52 1–33 (2019)
- Hao, S., Li, D., Halfond, W.G.J., Govindan, R.: *Estimating mobile application energy consumption using program analysis*. In: 2013 35th International Conference on Software Engineering (ICSE). pp. 92–101. (2013)
- Hindle, A., Wilson, A., Rasmussen, K., Barlow, E.J., Campbell, H.V., Romansky, S.: *GreenMiner: a hardware based mining software repositories software energy consumption framework*. In: MSR 2014. (2014)
- IBM Cloud Education: *SOA*, <https://www.ibm.com/cloud/learn/soa>, (2019)
- Ibrahim, N.: *Ranking Energy-Aware Services*. In: 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity). pp. 575–580. (2015)
- ISO/IEC 18384:2016, *Information technology — Reference Architecture for Service Oriented Architecture (SOA): Reference Architecture for SOA Solutions*, <https://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>, (2016)

- Manotas, I., Pollock, L., Clause, J.: *SEEDS: a software engineer's energy-optimization decision support framework*. In: Proceedings of the 36th International Conference on Software Engineering. pp. 503–514. Association for Computing Machinery, New York, NY, USA (2014)
- Niknejad, N., Ismail, W., Ghani, I., Nazari, B., Bahari, M., Hussin, A.R.B.C.: *Understanding Service-Oriented Architecture (SOA): A systematic literature review and directions for further investigation*. Inf. Syst. 91 101491 (2020)
- O'Brien, L., Bass, L., Merson, P.F.: *Quality Attributes and Service-Oriented Architectures*. (2005)
- Oliner, A., Iyer, A., Stoica, I., Lagerspetz, E., Tarkoma, S.: *Carat: collaborative energy diagnosis for mobile devices*. (2013)
- Oracle: *Software Engineering in an SOA Environment*, <https://www.oracle.com/technetwork/topics/entarch/oracle-pg-soa-sw-engineering-r3-2-1561707.pdf>, (2012)
- Patgiri, R., Ahmed, A.: *Big Data: The V's of the Game Changer Paradigm*. December 12 (2016)
- Pathak, A., Hu, Y.C., Zhang, M.: *Where is the energy spent inside my app? fine grained energy accounting on smartphones with Eprof*. In: Proceedings of the 7th ACM european conference on Computer Systems. pp. 29–42. Association for Computing Machinery, New York, NY, USA (2012)
- Peltonen, E., Lagerspetz, E., Nurmi, P., Tarkoma, S.: *Energy modeling of system settings: A crowdsourced approach*. In: 2015 IEEE International Conference on Pervasive Computing and Communications (PerCom). pp. 37–45. (2015)
- Pressman, R.S., Maxim, B.: *Software Engineering: A Practitioner's Approach*. , New York, NY (2014)
- Sahin, C., Cayci, F., Gutiérrez, I.L.M., Clause, J., Kiamilev, F., Pollock, L., Winbladh, K.: *Initial explorations on design pattern energy usage*. In: 2012 First International Workshop on Green and Sustainable Software (GREENS). pp. 55–61. (2012)
- Statista, *Cell phone sales worldwide 2007-2020*, <https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/>
- Wang, J., Feng, L., Xue, W., Song, Z.: *A survey on energy-efficient data management*. ACM SIGMOD Rec. 40 (2), 17–23 (2011)

Enrichir les traces GPS des visiteurs : enjeux et propositions

Cécile Cayère¹, Jérémy Richard², Mélanie Mondo³

1. La Rochelle Université
L3i
La Rochelle, France
cecile.cayere1@univ-lr.fr

2. La Rochelle Université
L3i
La Rochelle, France
jeremy.richard2@univ-lr.fr

3. La Rochelle Université
LIENSs
La Rochelle, France
melanie.mondo1@univ-lr.fr

RÉSUMÉ. À partir des traces GPS retraçant les itinéraires de touristes volontaires associées à leurs entretiens semi-directifs réalisés a posteriori, cet article a pour objectif de proposer une méthode d'enrichissement de ces traces de mobilité afin de comprendre plus finement les pratiques touristiques au sein d'une ville (La Rochelle). Cet enrichissement s'appuie sur la création de séquences d'épisodes sémantiques basée sur les entretiens ainsi que sur d'autres données contextuelles issues de l'Open Data. Chaque séquence qualifie la trace de mobilité avec un type de données particulier et s'organise en niveaux de précision.

ABSTRACT. Using GPS tracks tracing the itineraries of volunteer tourists associated with their semi-structured interviews, this article aims to propose a method of enriching these mobility tracks in order to understand more finely the tourist practices within a city (La Rochelle). This enrichment relies on the creation of semantic episode sequences based on the interviews as well as on other contextual data from the Open Data. Each sequence qualifies the mobility track with a specific type of data and is organized in levels of precision.

MOTS-CLÉS : trace numérique, trajectoire sémantique, pratique touristique

KEYWORDS: digital footprint, semantic trajectory, touristic practice

1. Introduction

La multiplication de traces de mobilité liées au développement de technologies mobiles entraîne de nombreuses promesses pour l'étude des pratiques touristiques en ville. Par leur potentielle finesse spatiale et temporelle, ces données permettraient de venir combler un besoin d'observation du territoire touristique pour les gestionnaires (Fournier, Jacquot, 2014).

Partant de ce constat, le projet régional et pluridisciplinaire DA3T¹ vise à proposer un dispositif d'analyse des traces numériques de mobilité afin d'améliorer la gestion et la valorisation des villes touristiques en Nouvelle-Aquitaine. Dans ce but, une enquête a été menée durant l'été 2020 sur les pratiques touristiques des visiteurs à La Rochelle où deux types de données ont été récoltées : un suivi GPS des visiteurs via leur téléphone portable et des entretiens semi-directifs. Travaillant conjointement entre informaticiens et géographes, nous avons développé un modèle d'enrichissement des traces GPS récoltées avec les données issues des entretiens et des données contextuelles issues de l'Open Data pour une analyse fine des pratiques touristiques.

Notre objectif est de concevoir un modèle de représentation de trajectoires sémantiques capable d'intégrer d'une part (1) les traces de mobilité de visiteurs et d'autre part, (2) des données d'enrichissement provenant soit des entretiens semi-directifs menés avec ces mêmes visiteurs, soit de l'Open Data. Nous utilisons la notion de séquences d'épisodes pour réaliser cet enrichissement.

Après avoir exposé quelques travaux connexes (section 2), nous présentons notre collecte de données (section 3) puis notre modèle d'enrichissement (section 4). Nous terminons par une mise en application de ce modèle sur une trace GPS d'un visiteur de La Rochelle (section 5).

2. Travaux connexes

La trace de mobilité permet de représenter le déplacement physique d'un objet mobile et est captée grâce à une technologie de capture (p. ex. GPS, caméra, IPS, etc.). Si le déplacement est par nature un phénomène continu dans l'espace et dans le temps, sa trace numérique est discrète en raison des limitations physiques des appareils de capture, de traitement et de stockage (Laube *et al.*, 2005). Une trace de mobilité consiste en une suite de positions spatio-temporelles où une position spatio-temporelle p est un tuple tel que : $p = (x, y, t)$ avec x et y les coordonnées de l'objet mobile à l'instant t (Parent *et al.*, 2013).

Les traces de mobilité GPS sont régulièrement utilisées pour l'étude du comportement des touristes (Corre *et al.*, 2012 ; Beeco *et al.*, 2013). Cependant, si leur intérêt réside dans leur potentiel pour observer l'emprise spatio-temporelle de la pratique touristique, elles ne renseignent pas sur l'intention des touristes,

1. Lien vers le site officiel : <http://da3t.iutbayonne.univ-pau.fr/>

sur les raisons sous-jacentes ou même sur le contenu de la pratique, contrairement aux informations qualitatives issues des entretiens semi-directifs. Quant aux données contextuelles issues de l'Open Data, elles permettent de contextualiser la trace et donc aident également à analyser finement le déplacement. Dans les travaux de (Skoumas *et al.*, 2015), données quantitatives et informations qualitatives sont clairement définies dans un contexte de recommandation d'itinéraire. Les données quantitatives sont des données fondées sur les coordonnées géographiques (c.-à-d. la longitude et la latitude). Lorsqu'un individu est amené à parler de son itinéraire, il préfère utiliser des informations qualitatives comme des toponymes (e.g. Tour de la Lanterne, Mairie de La Rochelle, etc.) et des relations spatiales (e.g. à côté, devant, derrière, etc.) les liant entre eux.

En informatique, le concept de trajectoire représente la sous-partie de la trace de mobilité qui a un intérêt pour l'application en question (Parent *et al.*, 2013). Dans le contexte de cet article, la trace entière est d'intérêt donc nous ne différencierons pas les deux concepts. Cette trajectoire peut être enrichie avec des données additionnelles provenant de sources diverses (p. ex. Open Data, entretiens, avis d'expert, etc.) afin de faciliter l'analyse des déplacements. La trajectoire brute ainsi enrichie devient une trajectoire sémantique (Parent *et al.*, 2013).

L'annotation est un processus d'enrichissement qui lie une donnée contextuelle à une trajectoire, à un segment de la trajectoire ou à une position de la trajectoire (Parent *et al.*, 2013). Un épisode est le segment de trajectoire maximal qui répond à un certain prédicat (Fileto *et al.*, 2013) (relatif aux annotations ou pas).

La représentation d'une trajectoire sémantique a la particularité de combiner les dimensions temporelle, thématique et spatiale. (Spaccapietra *et al.*, 2008) représentaient dès 2008 une trajectoire comme une séquence d'épisodes d'arrêts et de déplacements (*stops and moves*). Cette représentation de trajectoire sous forme de séquences d'épisodes est commune dans l'état de l'art. Le modèle APM (*Activity, POI and Move model*) (Moreau *et al.*, 2019), représente une trajectoire sémantique sous la forme d'une séquence d'épisodes. Un épisode est formalisé comme un couple $(s, [l, u])$ où s est un symbole qui représente la donnée contextuelle et $[l, u]$ est un intervalle où l est l'instant de début et u l'instant de fin. Chaque symbole du modèle APM est décrit par un concept dans une ontologie du domaine.

Nous proposons dans cet article, une façon de structurer des données d'enrichissement par le biais d'une séquence d'épisodes. Cette séquence nous permettra ainsi de compléter ou de détailler une trace GPS et de la visualiser.

3. Collecte de données

Afin de participer à notre enquête, les visiteurs installent sur leur téléphone mobile une application mobile², développée dans le cadre du projet, permettant d'enregistrer leurs déplacements. Le visiteur choisit la durée d'acquisition des données au moment de l'installation mais peut toujours décider d'interrompre la capture pendant cette période d'acquisition. Cela laisse au visiteur le contrôle sur ce qu'il décide ou non de transmettre. L'application enregistre à intervalles de temps réguliers la position spatio-temporelle p du mobile telle que $p = (x, y, t, D)$ avec x et y correspondant aux coordonnées géographiques de l'objet mobile (c.-à-d. la longitude et la latitude) au temps t de la capture et D étant un ensemble de données complémentaires brutes caractérisant le déplacement (p. ex. vitesse, orientation, précision de la capture, etc.). L'ensemble de ces captures p d'un visiteur forme sa trace brute qui représente son déplacement sur une certaine période temps. En complément des données de géolocalisation, le visiteur renseigne également ses dates de séjour, avec qui il voyage, son degré de connaissance de La Rochelle (première visite ou pas, déjà vécu ici, etc.) ainsi que son moyen d'arrivée. Ces informations permettent une première caractérisation des traces.

La qualité des données récoltées dépend beaucoup du matériel de capture. Les visiteurs participant à notre enquête possédaient des appareils mobiles différents, aux versions de systèmes d'exploitation différentes n'assurant pas une qualité constante ou homogène de la donnée récupérée. L'entretien permet ainsi également de valider la précision de la trace, afin de vérifier ce qui relève d'un manque de précision de l'application.

Nous cherchons à donner du sens aux traces GPS récoltées. Pour cela, nous allons utiliser deux type de données d'enrichissement : (1) des données contextuelles obtenues grâce à divers webservices de l'Open Data (p. ex. météo, événements, informations de fréquentation, etc.) et (2) les informations issues de l'entretien d'élucidation. La première permet de renseigner sur des faits globaux sur le contexte du déplacement qui ne seront pas forcément mentionnés dans l'entretien tandis que la seconde renseigne directement sur la pratique du producteur de la trace lui-même.

Dans cet article, nous utiliserons deux types de données contextuelles issues de l'Open Data. Premièrement, nous utiliserons des données de météo qui peuvent aider à répondre à certains questionnements concernant le déplacement d'une personne (p. ex. le visiteur est resté un long moment sous ces arcades et il pleuvait). Deuxièmement, nous utiliserons le découpage de la ville en quartier qui permet de comparer la vision du découpage de la ville par les gestionnaires avec celle du visiteur (p. ex. le visiteur dit qu'il est au centre-ville alors que selon le

2. Lien vers l'application Geoluciole : https://play.google.com/store/apps/details?id=fr.univ_lr.geoluciole

découpage de la ville par les gestionnaires le situe dans le quartier étudiant). Concernant l'entretien, nous avons choisi de nous inspirer des entretiens à base de *photo elicitation interview*, une "méthode d'enquête où l'entretien est mené sur la base d'un support photographique considéré comme susceptible de provoquer ou susciter des réactions verbales et émotionnelles chez la personne interviewée" (Bigando, 2013) en utilisant comme support visuel la carte des traces brutes du visiteur. Ainsi, la visualisation des traces GPS vient alimenter le discours de l'enquêté, et ce même discours permet d'alimenter la trace GPS, qui une fois enrichie et analysée permet d'interroger sous un nouvel angle l'entretien.

4. Modèle d'enrichissement

Le modèle d'enrichissement proposé permet de coupler les données quantitatives que sont les traces de mobilité GPS avec des données d'enrichissement contextuelles (c.-à-d. Open Data) ou qualitatives (c.-à-d. entretien d'élicitation) afin d'ajouter de la sémantique sur des données a priori brutes.

Introduit par (Allen, 1983), la notion d'intervalles temporels permet de représenter des choses simples telles qu'une journée type du quotidien, sous forme de séquences d'intervalles temporels (p. ex. de 8h00 à 8h30 petit-déjeuner, de 8h30 à 12h00 travail) formalisées comme suit pour une séquence d'intervalle S , $S = (s_i, [l_i, u_i])_{i \in \mathbb{N}}$ (Guyet, Quiniou, 2011), où s est une information sémantique et $[l, u]$, où $l, u \in \mathbb{R}, l < u$ sont des dates. Nous appellerons un élément de cette séquence ordonnée est un épisode.

Ce modèle vise à représenter de manière la plus exhaustive toutes les données d'enrichissement dont nous disposons. Cet objectif nous pousse à vouloir caractériser certains éléments d'une séquence, voire même dans certains cas, les généraliser en un seul et même épisode plus large. P. ex. on peut vouloir considérer une suite d'épisodes A : $A = (\text{vélo}, [8h00, 9h30]), (\text{marche}, [9h30, 11h00]), (\text{repas}, [11h00, 13h00])$ comme étant : $(\text{circuit touristique}, [8h00, 13h00])$. Ainsi, nous considérons que certains épisodes peuvent être composés de plus courts épisodes. Nous appelons ce genre d'abstraction, un niveau. Fondamentalement, on considère l'épisode *circuit touristique*, comme étant un épisode de niveau 1, et quant à la suite d'épisodes A , nous la qualifions de niveau 2. Plus nous entrons dans les détails de ce que compose un épisode, plus nous augmentons la valeur du niveau. Par conséquent, chaque élément $s_{i,n}$ de la séquence S_n peut contenir lui-même une séquence S_{n+1} d'épisode, où $n \in \mathbb{N}$ est le niveau de cette séquence. Nous formalisons comme suit : $(s_{n+1,i}, [l_{n+1,i}, u_{n+1,i}])_{i \in \mathbb{N}} \sqsubset (s_{n,j}, [l_{n,j}, u_{n,j}])$. Ce modèle a donné lieu à la création d'un diagramme de classe dont la dernière version est disponible en ligne³.

3. Lien vers le diagramme de classe : https://github.com/Caecilia/thesis/blob/master/doc/class_diagram.png

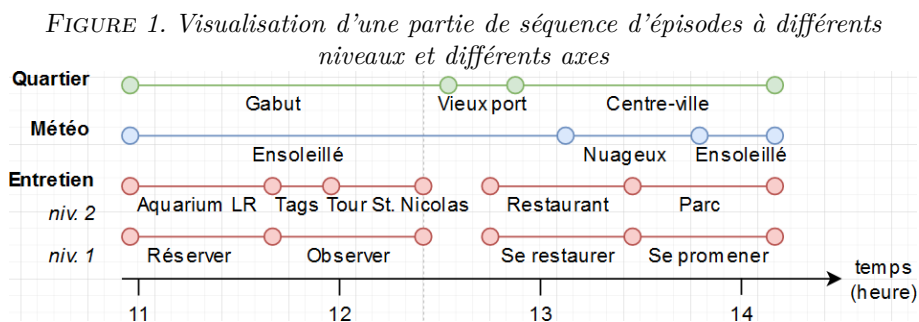
5. Application aux données collectées : enrichissement de traces GPS d'un visiteur à La Rochelle

Afin d'éprouver notre modèle, nous nous sommes appuyés sur des données récoltées à l'été 2020. Nous disposons de la dizaine de traces GPS récoltées depuis les téléphones mobiles des visiteurs, des entretiens semi-directifs associés à ces traces, des données de météo pour cette période et du découpage administratif des quartiers de la ville. Nous prenons l'exemple d'une personne visitant la Rochelle pour la journée, pour la première fois, en famille et arrivée en voiture. Ces données sont ainsi stockées dans notre modèle.

Voici un extrait de l'entretien que nous transformerons en séquence d'épisodes dans la suite de l'expérimentation : *"Je connais pas le nom du quartier mais pas très loin de l'office du tourisme, il y a un endroit où les gens font des tags sur les murs pas loin du port. On s'est bien arrêté une bonne dizaine de minutes pour regarder ce qu'ils faisaient. Après on s'est dirigé vers la tour Saint-Nicolas, on s'est arrêté un petit moment là aussi à regarder. [...] On n'a pas visité les tours. Puis on a passé du temps sur les écriteaux qui expliquaient ce qui s'est passé dans la ville avec Richelieu, etc. [...]"*.

5.1. Instanciation du modèle

Dans l'extrait présenté, plusieurs épisodes sont évoqués mais avec peu d'informations temporelles précises. Nous utilisons donc tant le discours que les traces associées pour retrouver le découpage temporel des épisodes racontés. La figure 1 montre plusieurs séquences d'épisodes visant à enrichir la trace dont celle correspondant à un découpage du discours de l'enquêté sur deux niveaux. Les noms des épisodes correspondent à des concepts de vocabulaires spécifiques mais leur description sort du cadre de ce travail.

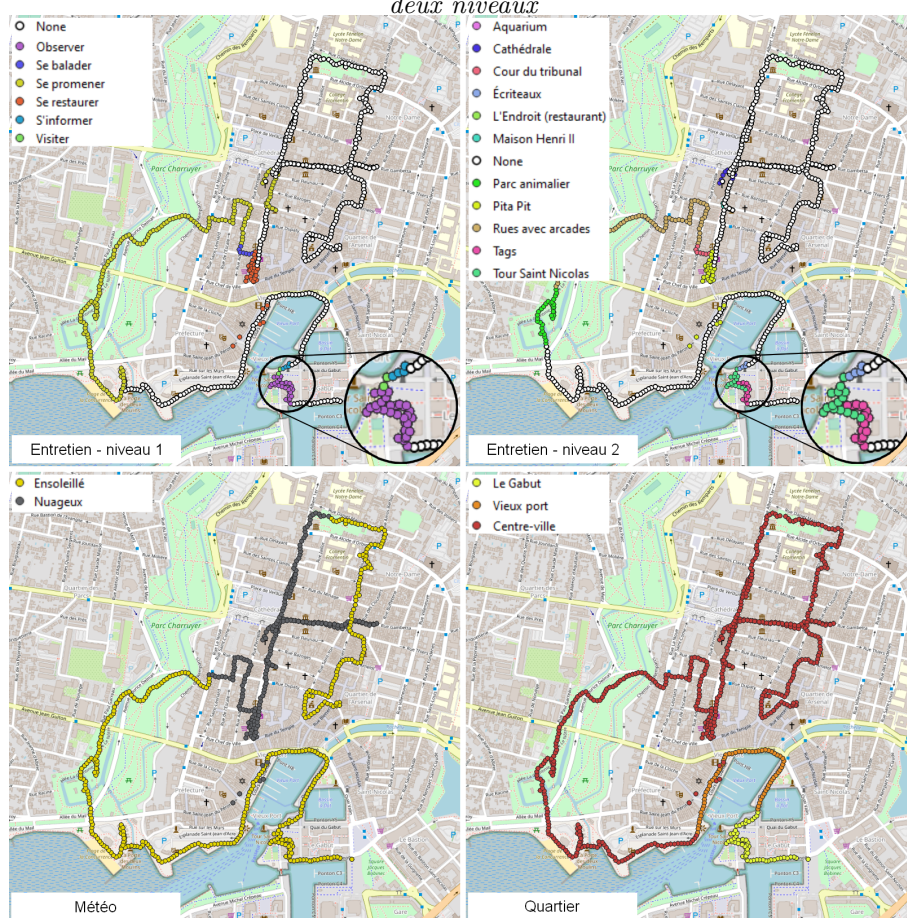


Ainsi, nous avons caractérisé différents épisodes basés sur les données d'enrichissement, permettant d'enrichir la trace selon les différents axes et niveaux.

5.2. Visualisation des séquences d'épisodes

L'annotation ainsi complétée depuis l'extrait d'entretien, il est possible de visualiser la trace enrichie sémantiquement de façon cartographique. Dans la figure 2, nous représentons ainsi notre parcours touristique selon les trois axes d'enrichissement vus précédemment.

FIGURE 2. Représentation des trois axes d'enrichissement dont l'un est sur deux niveaux



L'axe d'enrichissement correspondant à l'entretien est construit sur deux niveaux, les séquences du second précisant les épisodes du premier. Le premier niveau permet de visualiser l'action principale de la pratique touristique alors que le second répertorie les lieux évoqués pendant l'entretien. Ainsi, *Tags* (c.f. figure 2, Entretien - niveau 2) correspondant au toponyme usuel la "Friche du Gabut", une ancienne friche remarquable dans le quartier du Gabut (c.f. figure

2, Quartier).

Si la carte nous permet de visualiser les pratiques associées à la trace, elle met également en avant le blanc du discours. Ainsi, la balade autour du Vieux-Port (c.f. figure 2, Quartier) n'est jamais évoquée alors qu'il s'agit d'un lieu emblématique de La Rochelle : cette absence d'informations se retrouve dans la modalité *None* de nos légendes. C'est ainsi que l'enrichissement via des données contextuelles issues de l'Open Data comme les quartiers de La Rochelle peut s'avérer utile. En effet, nous pouvons à l'aide de cet enrichissement que le visiteur à longé le quartier du *Vieux port*. À l'inverse, des informations contenues dans l'entretien ne sont pas visibles dans cette cartographie : le début du séjour où l'application n'était pas encore installée, mais également la fin de journée suite à un arrêt imprévu de l'application. Si la suite du séjour a été évoquée en entretien, il n'est pour le moment pas visible sur cette carte.

6. Conclusion et perspectives

Partant de la problématique de l'enrichissement sémantique des traces GPS des visiteurs, nous cherchions à éprouver un modèle informatique intégrant à la fois des informations qualitatives issues d'un entretien semi-directif et des données contextuelles issues de l'Open Data et les transformant en séquence d'épisodes. Ces séquences permettent d'apporter une compréhension fine des pratiques touristiques, d'abord à l'échelle individuelle et à terme, de façon agrégée. Ce travail exploratoire nous a permis de valider l'intérêt d'une telle méthode.

Par ailleurs, le découpage par séquence temporelle montre également ses limites. En effet, pour le moment, les épisodes d'une séquence s'enchaînent chronologiquement. Or, dans la pratique touristique, les activités peuvent régulièrement se superposer totalement ou partiellement (p. ex. le visiteur mange un sandwich en visitant le centre-ville). Une piste pour résoudre ce problème est la notion d'intervalles flous (Freksa, 1992) qui fera l'objet de travaux futurs. Cela permettrait également de prendre en compte aussi bien une information temporelle précise (p. ex. le visiteur est allé à la plage de 14h30 à 16h00) qu'une information plus relative (p. ex. le visiteur est allé à la plage après être allé à l'Aquarium) et ainsi coller plus facilement au discours de l'enquêté.

En perspective, ce modèle de représentation est tout à fait adapté pour effectuer de la fouille de motif (ie. *pattern mining*), afin de révéler les pratiques et les contextes courants chez les visiteurs. La fouille de motif dans les séquences d'intervalles étant un sujet déjà exploré dans l'état de l'art de ce domaine, il serait intéressant d'observer ce que la notion de niveau peut apporter en terme de résultats. Cela pourrait par ailleurs nécessiter de nous intéresser au traitement automatique du langage afin d'appliquer ce modèle à des corpus de données plus importants. Cela permettrait alors de proposer une analyse plus générique des pratiques touristiques.

Remerciements

Cet article a été écrit dans le cadre de trois thèses (informatique et géographie), dans le projet DA3T, financées par la région Nouvelle-Aquitaine et la société Berger-Levrault.

Bibliographie

- Allen J. F. (1983). Maintaining knowledge about temporal intervals. *Commun. ACM*, vol. 26, n° 11, p. 832 – 843.
- Beeco J. A., Huang W.-J., Hallo J. C., Norman W. C., McGehee N. G., McGee J. et al. (2013). GPS Tracking of Travel Routes of Wanderers and Planners. *Tourism Geographies*, vol. 15, n° 3, p. 551 – 573.
- Bigando E. (2013). De l'usage de la photo elicitation interview pour appréhender les paysages du quotidien : retour sur une méthode productrice d'une réflexivité habitante. *Cybergeo : European Journal of Geography*.
- Corre N. L., Berre S. L., Brigand L., Peuziat I. (2012). Comment étudier et suivre la fréquentation dans les espaces littoraux, marins et insulaires? De l'état de l'art à une vision prospective de la recherche. *EchoGéo*, n° 19.
- Fileto R., Krüger M., Pelekis N., Theodoridis Y., Renso C. (2013). Baquara: A Holistic Ontological Framework for Movement Analysis Using Linked Data. In W. Ng, V. C. Storey, J. C. Trujillo (Eds.), *Conceptual Modeling*, p. 342 – 355.
- Fournier C., Jacquot S. (2014). Les traces numériques des touristes. Un renouvellement de l'observation touristique? *Espaces Tourismes et loisirs*, n° 316, p. 66 – 71.
- Freksa C. (1992). Temporal reasoning based on semi-intervals. *Artificial Intelligence*, vol. 54, n° 1-2, p. 199 – 227.
- Guyet T., Quiniou R. (2011). Extracting temporal patterns from interval-based sequences. In *International Joint Conference on Artificial Intelligence (IJCAI)*.
- Laube P., Imfeld S., Weibel R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, vol. 19, n° 6, p. 639 – 668.
- Moreau C., Devogele T., Etienne L. (2019). Calcul de similarité sémantique entre trajectoires. *Revue Internationale de Géomatique*, vol. 29, n° 1, p. 107 – 127.
- Parent C., Spaccapietra S., Renso C., Andrienko G. L., Andrienko N. V., Bogorny V. et al. (2013). Semantic trajectories modeling and analysis. *CSUR*.
- Skoumas G., Schmid K. A., Jossé G., Schubert M., Nascimento M. A., Züfle A. et al. (2015). Knowledge-Enriched Route Computation. In C. Claramunt et al. (Eds.), *Advances in Spatial and Temporal Databases*, vol. 9239, p. 157 – 176. Cham, Springer International Publishing.
- Spaccapietra S., Parent C., Damiani M., Antonio de Macedo J., Porto F., Vangenot C. (2008). A conceptual view on trajectories. *Data Knowledge Engineering*, vol. 65, n° 1, p. 126 – 146.

Le projet Pulse : vers la supervision des échanges dans un système IoT

Jannik Laval¹, Boubou Thiam Niang¹, Imene Ghzaïel¹,
Kenza Riahi¹, Baudouin Dafflon², Giacomo Kahn¹,
Yacine Ouzrout¹

Univ Lyon, Univ Lumière Lyon 2, INSA Lyon, Université Claude Bernard Lyon 1,
DISP, EA4570, 69676 Bron, France
prenom.nom@univ-lyon2.fr

ABSTRACT. This article describes our work on monitoring data exchanges in an IoT information system. We address the collection, modelling and processing of data in order to analyse the evolution of the system architecture and identify the problems that may arise.

RÉSUMÉ. Cet article décrit nos travaux dans le domaine de la supervision des échanges de données dans un système d'information IoT. Nous abordons la collecte, la modélisation et le traitement des données afin d'analyser l'évolution de l'architecture du système et d'identifier les problèmes qui peuvent survenir.

KEYWORDS: Data Monitoring, IoT System, Dynamic analysis of IoT, Interoperability Assessment.

MOTS-CLÉS : Surveillance des données, système IoT, analyse dynamique de l'IoT, évaluation de l'interopérabilité.

1. Introduction

Les architectures orientées événements (EDA) (Sriraman, Radhakrishnan, 2005) permettent un couplage lâche, mais limitent la capacité de compréhension du comportement global de l'architecture. Cette limitation est due à la multiplicité des échanges de données. Par conséquent, il est nécessaire de mettre en place des systèmes de surveillance et d'analyse pour mieux comprendre le fonctionnement et l'évolution en temps réel du système d'échanges (Brand, Giese, 2019). Il existe, à ce jour, des plateformes proposant des consoles de supervision. Cependant, la plupart de ces solutions traitent principalement des informations de bas niveau, telles que la fréquence des messages échangés ou l'utilisation de la mémoire. Le but du travail de recherche présenté dans cet article est de proposer une solution permettant de collecter les informations et les messages provenant des systèmes IoT afin de les analyser. La supervision des systèmes IoT implique d'intégrer l'aspect dynamique de leur architecture

et la gestion de l'imprévu, tels que les problèmes de connexion ou de sécurité. La collecte d'informations provenant du système d'une part et des messages d'autre part est une première étape pour analyser un système IoT en fonctionnement.

2. Le framework Pulse

Un système de surveillance devrait fournir des moyens permettant de maintenir un niveau d'interopérabilité conforme aux exigences d'interopérabilité (Mallek *et al.*, 2010). Dans ce sens, nous proposons une approche permettant d'effectuer une supervision avancée en combinant les informations de l'architecture et les métadonnées des messages échangés. Ainsi, cela permet de produire les indicateurs nécessaires à la détermination des actions de maintenance à effectuer sur le système. Pour cela, nous proposons le *framework* Pulse (Figure 1) qui est composé de quatre niveaux : l'importation des données, la gestion du temps et des modifications de l'architecture, la persistance et l'analyse des données.

Un prototype du *framework* est déjà opérationnel ¹ et permet de visualiser certains comportements du système. Il facilite ainsi l'interprétation du statut de chaque message et fournit des indicateurs pour l'analyse de l'interopérabilité. Le prototype est implémenté sur Moose, une plateforme d'analyse de données open source (Demeyer *et al.*, 2001; Ducasse *et al.*, 2009), dans laquelle le métamodèle Pulse est intégré.

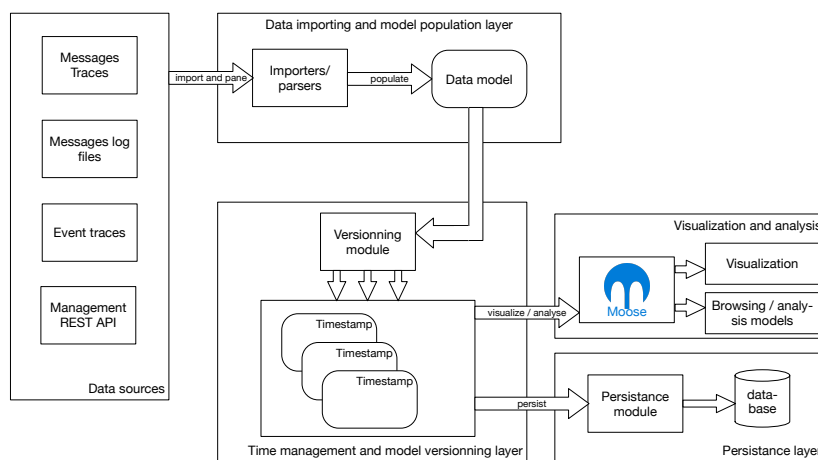


Figure 1. Le framework Pulse.

Le niveau "importation" collecte les métadonnées des messages transitant dans le système IoT et reconstitue l'architecture. Ces métadonnées permettent de suivre les messages échangés, d'identifier les entités actives, et d'améliorer la maintenabilité grâce à la détection des exceptions. Les métadonnées collectées sont organisées

1. <https://github.com/janniklaval/pulse>

pour être traitées de manière cohérente dans un métamodèle générique et extensible, présenté en Section 3.

Le niveau “gestion du temps” permet de traiter trois aspects du système : une représentation statique, considérant notamment les entités et les canaux par lesquels transitent les messages ; une représentation dynamique, qui concerne le cycle de vie des messages (connexions, messages et toute entité volatile) ; et la représentation du cycle de vie de l’architecture, où les entités sont créées, modifiées et supprimées. En conséquence, la structure Pulse intègre des fonctionnalités de modélisation dynamique de l’architecture en incluant un historique contenant les dates de création et de suppression des composants. Cette gestion du temps est implémentée avec Orion (Laval *et al.*, 2011), un outil interactif de prototypage pour la réingénierie.

Le niveau “analyse” permet de définir un ensemble d’indicateurs qui évaluent l’état du système au cours du temps ou à un instant donné pour visualiser et analyser les modifications et leur impact. Ces deux fonctionnalités nous permettent non seulement de détecter les problèmes au moment où ils se produisent, mais aussi d’identifier la source potentielle d’un problème en revenant à des états antérieurs.

3. Le Métamodèle Pulse

L’objectif du métamodèle Pulse est de représenter les trois aspects du système : la structure statique, le cycle de vie des messages et le cycle de vie de l’architecture.

Plusieurs protocoles peuvent être utilisés dans un même système d’information, un des objectifs de ce métamodèle est d’être suffisamment générique pour prendre en compte différents protocoles, et d’être extensible. Ce métamodèle est capable de prendre en charge les protocoles AMQP, MQTT, KAFKA et CoAP (Amokrane *et al.*, 2018; Laval *et al.*, 2020; Laval, 2021). Nous travaillons actuellement à son adaptation au *middleware* ROS (*Robot Operating System*). La Figure 2 illustre la structure actuelle du métamodèle.

Il rassemble des informations provenant de plusieurs sources. Si une console de gestion existe, les traces de messages fournies par cette console, cela permet d’identifier pour chaque message :

- Le nœud par lequel il transite, ainsi que les informations de connexion et d’hôte.
- L’échange dans lequel il a été publié ou consommé, les files d’attente vers lesquelles il est acheminé ou la file d’attente depuis laquelle il est consommé, ainsi que les clés de routage associées.
- L’utilisateur qui publie ou consomme le message.
- Son horodatage, son type (publié / reçu) et son mode de livraison (persistant ou non).

Les événements de création et de suppression de ressources, d’hôtes, d’utilisateurs et de permissions ; la création et la fermeture de connexions et de canaux et les tentative d’authentification de l’utilisateur peuvent également être récupérés par l’analyse des

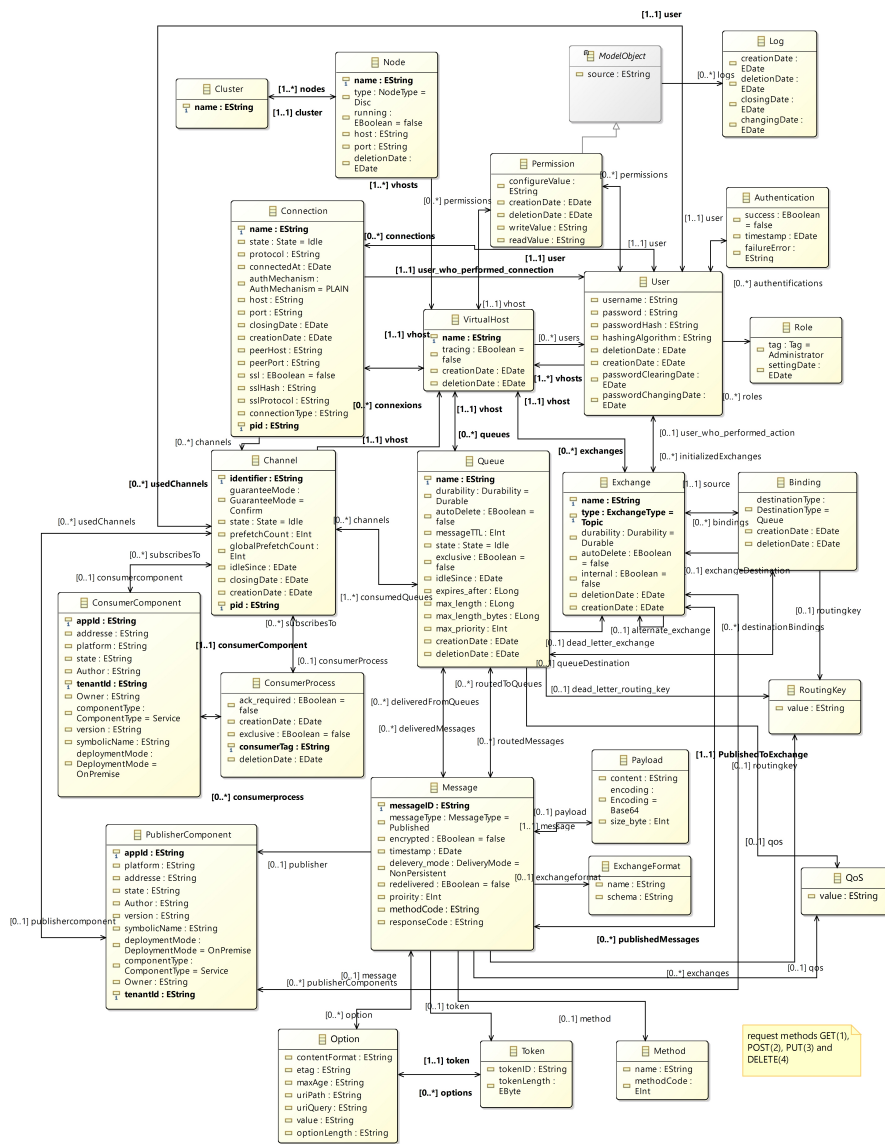


Figure 2. The Pulse Metamodel

traces. Les éléments contextuels sur les caractéristiques des applications communicantes fournis sont extraites des métadonnées des messages.

4. Défis et verrous

Le prototype présenté montre la faisabilité et l'intérêt de notre approche. Son implémentation permettant d'analyser des systèmes AMQP d'une entreprise partenaire nous encourage à l'étendre vers d'autres architectures orientées événements. Les perspectives sont de plusieurs ordres, énoncés ci-après.

- Aide à la conception : Le métamodèle Pulse peut être utilisé dans la phase de conception des systèmes. L'analyse des échanges de données peut permettre d'extraire les caractéristiques des connecteurs permettant aux nouveaux constituants de se connecter au système d'échange. Les connecteurs sont des entités de première classe de l'architecture qui décrivent la sémantique de communication entre des constituants hétérogènes (Kutsche *et al.*, 2008). Ils sont donc des constituants à part entières avec des fonctionnalités leur permettant d'établir la communication entre d'autres constituants du système, par exemple le filtrage des informations sensibles ou la traduction du format du message dans un autre format. Dans ce sens, nous comptons utiliser Pulse dans un projet de configuration et de génération automatique des connecteurs permettant d'intégrer une entité dans un système d'information. Dans cette proposition, un modèle du connecteur est généré à partir d'une configuration du *feature model* représentant l'ensemble des fonctionnalités d'un connecteur désiré. Ainsi en plus de l'aspect données, le métamodèle sera étendu pour tenir compte de la structure du *feature model* et des fonctionnalités que ce dernier représente. D'autre part, les caractéristiques de chaque connecteur varient en fonction du besoin de communication: mode synchrone ou asynchrone, type de transport point-à-point ou *publish/subscribe*, etc. De plus les connecteurs doivent être générés automatiquement. Il faudra s'assurer que cette extension sera applicable dans un contexte évolutif et ouvert en tenant en compte des fonctionnalités variables du connecteur.

- Vérification de la qualité des échanges : obtenir des mesures quantitatives sur l'échange de données en temps réel nous permet d'élaborer une approche d'évaluation automatique de la fiabilité du système. Nous travaillons à définir un ensemble d'indicateurs en prenant en compte les caractéristiques et l'architecture du système IoT et les mesures associées issues de la norme (ISO/IEC 30141, 2018). Nous ne gardons que les indicateurs les plus pertinents et mesurables pour les associer adéquatement aux problèmes d'interopérabilité. Différentes techniques d'analyse peuvent être adoptées pour identifier les causes potentielles des problèmes par exemple les arbres de décision (Ringwald *et al.*, 2006) ou les méthodes à base de modèles (Khan *et al.*, 2010). L'objectif est d'utiliser le *framework* Pulse pour quantifier les échanges en temps réel et appliquer des algorithmes permettant de contrôler l'état du système en fonctionnement, mais également de fiabiliser la phase de conception.

- Simulation et banc de tests : la simulation d'un système IoT et de ses échanges est intéressante pour rendre plus rapides la conception et la maintenance par les tests. La simulation peut être complète ou hybride en intégrant du matériel dans la boucle (Hardware in the Loop - HIL). Elle permet de tester des scénarios difficilement testable en situation réel, dangereux ou trop coûteux.

Notre objectif est de pouvoir définir des scénarios, contraintes et métriques afin de mettre en évidence les points de ruptures des algorithmes évalués. Lors de ces simulations, nous apportons un intérêt particulier à l'étude des anomalies temporaires ou définitives et à la réponse que le système adoptera pour y faire face. Pour assurer la capacité d'exploration du simulateur, un ensemble de fichiers de configurations a été standardisé et documenté. Les travaux actuels permettent d'instancier un ensemble de capteurs réels et/ou virtuels interagissant ensemble. Le format de données envoyées vers le système IoT a été standardisé pour tous les capteurs afin d'assurer une forte interopérabilité. Pour garder une cohérence entre les capteurs du banc de test et les capteurs virtuels, la génération des données produites par ces capteurs s'appuie sur des algorithmes d'apprentissage automatique profonds. Ces modèles permettent de générer des données exploitables semblables à des données provenant de capteurs réels.

– Sûreté de fonctionnement du système : l'analyse de fiabilité et de la sûreté de fonctionnement peut être complétée avec les données de suivi. Nous envisageons le développement de pronostics de défaillance (*Prognostics and Health Management*) du système IoT. Le pronostic de défaillance, bien connu dans le domaine de la mécanique (Atamuradov *et al.*, 2017), transforme un ensemble de données recueillies sur l'équipement surveillé en indicateurs de santé du système, extrapolés dans le temps pour aider à sa maintenance. Nous envisageons de développer ce type de solutions sur les systèmes cyberphysiques. En couplant le *framework* Pulse avec des approches de pronostics de défaillance, nous espérons obtenir une gestion fine du cycle de vie des composants du système.

5. Conclusion

Les travaux menés dans l'analyse des données d'un système d'échanges de type EDA nous montrent que les systèmes IoT, bien que très flexibles, n'offrent pas de moyens de supervision avancés. Nos travaux s'attachent à aborder ces problèmes en considérant l'ensemble du cycle de vie du système (conception, analyse en fonctionnement, maintenance). Le *framework* Pulse et le métamodèle proposés ainsi que son prototype nous offre une base solide pour envisager d'aborder les défis présentés dans cet article. Chacun de ces défis est actuellement en cours d'investigation et lèvera sans doute de nouveaux verrous.

Dans cet article, nous n'avons pas considéré les verrous quant à l'exploitation de notre système. Notamment, l'un d'eux est que l'approche est actuellement centralisée. Or, nous pourrions envisager le *framework* comme un environnement de type système de systèmes. Ce changement d'architecture permettrait de passer à l'échelle grâce à une répartition de la charge de traitement des données. En même temps, des algorithmes de fusion de données ou de fusion de graphes devront être développés pour obtenir de cette supervision distribuée, une vision globale du système.

References

- Amokrane N., Laval J., Lanco P., Derras M., Moala N. (2018). Analysis of data exchanges, contribution to data interoperability assessment. In *2018 international conference on intelligent systems (is)*, p. 199-208.
- Atamuradov V., Medjaher K., Dersin P., Lamoureux B., Zerhouni N. (2017). Prognostics and health management for maintenance practitioners-review, implementation and tools evaluation. *International Journal of Prognostics and Health Management*, Vol. 8, No. 060, pp. 1–31.
- Brand T., Giese H. (2019). Generic adaptive monitoring based on executed architecture runtime model queries and events. In *2019 IEEE 13th international conference on self-adaptive and self-organizing systems (saso)*, p. 17-22.
- Demeyer S., Tichelaar S., Ducasse S. (2001). *FAMIX 2.1—the FAMOOS information exchange model*. Technical report, University of Bern.
- Ducasse S., Girba T., Kuhn A., Renggli L. (2009). Meta-environment and executable meta-language using smalltalk: an experience report. *Software & Systems Modeling*, Vol. 8, No. 1, pp. 5–19.
- Internet of Things (IoT) – Reference architecture*. International Standard. (2018). International Organization for Standardization.
- Khan M. M. H., Le H. K., LeMay M., Moinzadeh P., Wang L., Yang Y. *et al.* (2010). Diagnostic powertracing for sensor node failure analysis. In *Proceedings of the 9th ACM/IEEE international conference on information processing in sensor networks*, pp. 117–128.
- Kutsche R., Milanovic N., Bauhoff G., Baum T., Carlsburg M., Kumpe D. *et al.* (2008). Bizycle: Model-based interoperability platform for software and data integration. *Proceedings of the MDTPI at ECMDA*, Vol. 430.
- Laval J. (2021). The pulse project: a framework for supervising data exchanges in an IoT system. *Automation, Robotics & Communications for Industry 4.0*, pp. 74.
- Laval J., Amokrane N., Derras M., Moalla N. (2020, Nov). Analysis of data exchange among heterogeneous IoT systems. In *10th international conference on interoperability for enterprise systems and applications, Tarbes*.
- Laval J., Denier S., Ducasse S., Falleri J.-R. (2011). Supporting simultaneous versions for software evolution assessment. *Science of Computer Programming*, Vol. 76, No. 12, pp. 1177–1193.
- Mallek S., Daclin N., Chapurlat V. (2010). Towards a conceptualisation of interoperability requirements. In K. Popplewell, J. Harding, R. Poler, R. Chalmers (Eds.), *Enterprise interoperability iv*, pp. 439–448. London, Springer London.
- Ringwald M., Römer K., Vialetti A. (2006). *Snif: Sensor network inspection framework. department of computer science*. Technical report. ETH Zurich, Technical Report, 535.
- Sriraman B., Radhakrishnan R. (2005). Event driven architecture augmenting service oriented architectures. *Report of Unisys and Sun Microsystems*.

Atelier

SimpleText : Simplification et Vulgarisation des Textes Scientifiques

Préface

Liana ERMAKOVA¹, Josiane MOTHE², Eric SANJUAN³

1. HCTI - EA 4249, Université de Bretagne Occidentale
20 Rue Duquesne, 29200 Brest, France
liana.ermakova@univ-brest.fr

2. INSPE, Université de Toulouse, IRIT, UMR5505 CNRS
Toulouse, France
josiane.mothe@irit.fr

3. Avignon Université, LIA, France
Avignon, France
eric.sanjuan@univ-avignon.fr

1. Introduction

Les systèmes modernes d'accès à l'information promettent de donner aux citoyens un accès direct à des informations clés provenant de sources primaires qui font autorité. La littérature scientifique est également concernée mais elle est en réalité difficilement accessible aux non-experts en raison de sa complexité langagière, la structure, longueur, etc. des documents scientifiques et du manque d'acculturation scientifique des lecteurs en général.

La teneur du débat scientifique qui procède par confrontation d'une multiplicité d'études avant de parvenir à un consensus est aussi source de complexité. Les décisions individuelles ou politiques sont potentiellement impactées par une méconnaissance de l'ensemble des travaux et débats scientifiques. Les scientifiques peuvent également être confrontés à une difficulté de lecture lorsqu'ils s'intéressent aux documents scientifiques de disciplines dans lesquelles ils ne sont pas experts. Les résultats contradictoires à l'intérieur même d'une discipline sont difficilement appréhendables par des non-spécialistes. Quid alors des résultats potentiellement contradictoires dans différentes disciplines ?

La simplification de textes se donne pour objectif de réduire ces obstacles. Ainsi l'atelier SimpleText aborde les opportunités et les défis des approches de simplification de textes scientifiques pour améliorer l'accès à l'information et l'acculturation scientifique. SimpleText a pour objectif de faire un pas vers une recherche véritablement ouverte, accessible et compréhensible pour tous, contribuer à contrer les fausses nouvelles fondées sur des résultats scientifiques ; ainsi que permettre une lecture plus rapide, ce qui peut également faciliter l'accès aux résultats scientifiques. Cela est particulièrement important compte tenu de l'explosion de la science ouverte et des preprints pendant la pandémie de la COVID-19 (e.g. SAGE Publishing¹ et Springer Nature² ont mis en open access les publications sur la COVID-19). Les textes simplifiés peuvent également avoir pour objet d'être plus accessibles aux locuteurs non natifs, aux jeunes lecteurs et aux personnes souffrant de troubles de la lecture. Ils peuvent permettre d'améliorer les applications du traitement du langage naturel, notamment celles de traduction automatique.

Ainsi, la simplification automatique des textes pourrait être utile dans divers domaines tels que la communication scientifique, le journalisme scientifique, la politique, la traduction assistée par ordinateur, la rédaction technique et l'éducation.

L'atelier SimpleText est un lieu de réflexion et de travail autour de ces questions. Cet atelier s'inscrit dans le cadre du Groupement de Recherche MaDICS³ (Masses de Données, Informations et Connaissances en Sciences).

L'atelier SimpleText s'appuie sur une communauté interdisciplinaire de chercheurs en traitement automatique de la langue, recherche d'information, linguistique, sociologie, journalisme scientifique et vulgarisation scientifique travaillant ensemble pour tenter de résoudre l'un des plus grands défis d'aujourd'hui.

2. Publications retenues

Cinq articles scientifiques qui couvrent les domaines de la simplification de textes médicaux, la didactique, la traduction et la relecture, l'intelligence artificielle et la rédaction technique, ont été retenus pour la publication dans les actes « Ateliers d'INFORSID - dessinons ensemble le futur des systèmes d'information » (dans l'ordre alphabétique) :

- Sílvia Araújo (Centre d'Études Humaines - CEHUM, Université du Minho, Braga, Portugal) et Radia Hannachi (HCTI EA-4249, Université de Bretagne Sud, Lorient, France)
Pour une démarche de communication multimodale de données scientifiques : de la recherche documentaire à l'infographie via le mind mapping

¹ <https://journals.sagepub.com/coronavirus>

² <https://www.springernature.com/fr/researchers/campaigns/coronavirus>

³ <https://www.madics.fr/ateliers/simpletext/>

- Rémi Cardon et Natalia Grabar (UMR 8163 STL – CNRS / Université de Lille, F-59000 Lille, France)
Recherche de phrases parallèles à partir de corpus comparables pour la simplification de textes médicaux en français (conférence invitée)
- Helen McCombie-Boudry (Bureau de Traduction de l'Université, Université de Brest, France)
Could automatic text simplification assist correction-revision of scientific texts written by non-native English speakers?
- John Rochford (Eunice Kennedy Shriver Center, University of Massachusetts Medical School, USA)
Developing Simple Web Text for People with Intellectual Disabilities and to Train Artificial Intelligence (conférence invitée)
- Mike Unwalla (TechScribe, Sheffield, UK)
Controlled language for text simplification: Concepts and implementation (présentation industrielle)

3. Le comité de programme

Les publications ont été relues par le comité de programme interdisciplinaire issu des domaines variés comme informatique, linguistique, épistémologie et histoire des sciences et des techniques. Nous voudrions ici remercier notre comité de programme:

- Ismail Badache, Univ. d'Aix-Marseille, Univ. de Toulon CNRS, LIS UMR 7020 (Marseille, France)
- Patrice Bellot, Univ. d'Aix-Marseille, Univ. de Toulon CNRS, LIS UMR 7020 (Marseille, France)
- Pavel Braslavski, Combinatorial Algebra Lab, Ural Federal University (Yekaterinburg, Russia)
- Adrian-Gabriel Chifu, Univ. d'Aix-Marseille, Univ. de Toulon CNRS, LIS UMR 7020 (Marseille, France)
- Liana Ermakova, HCTI - EA 4249, Université de Bretagne Occidentale (Brest, France)
- Hervé Ferrière, Centre François Viète d'épistémologie et d'histoire des sciences et des techniques, Université de Bretagne Occidentale (Brest, France)
- Jaap Kamps, Faculty of Humanities, University of Amsterdam (Amsterdam, Netherland)
- Élise Mathurin, HCTI - EA 4249, Université de Bretagne Occidentale (Brest, France)

- Josiane Mothe, INSPE, Université de Toulouse, IRIT, UMR5505 CNRS (Toulouse, France)
- Diana Nurbakova, LIRIS, Institut National des Sciences Appliquées de Lyon, (Lyon, France)
- Irina Ovchinnikova, Institute of Linguistics and Intercultural Communication, Sechenov University (Moscow, Russia)
- Michael Rinn, HCTI - EA 4249, Université de Bretagne Occidentale (Brest, France)
- Eric San-Juan, Laboratoire d'Informatique d'Avignon, Institut de technologie d' Avignon (Avignon, France)

Recherche de phrases parallèles à partir de corpus comparables pour la simplification de textes médicaux en français

Rémi Cardon¹, Natalia Grabar¹

UMR 8163 STL – CNRS / Université de Lille, F-59000 LILLE

{remi.cardon,natalia.grabar}@univ-lille.fr

RÉSUMÉ. La simplification automatique a pour objectif de produire une version de textes plus facile à comprendre à destination d'un public identifié. Nous nous intéressons à la simplification de textes médicaux. Le plus souvent, le lexique et les règles de simplification sont acquis à partir de corpus parallèles. Comme de tels corpus n'existent pas ou sont très rares en français, nous proposons des méthodes pour les construire à partir de corpus comparables. Notre méthode repose sur une étape de filtrage, destinée à ne garder que les meilleures phrases candidates à l'alignement, et une étape d'alignement considérée comme un problème de catégorisation. Il s'agit de décider si une paire de phrases est alignable ou non. Nous exploitons différents types de descripteurs (essentiellement basés sur le lexique et les corpus). Les résultats obtenus permettent de construire un corpus parallèle pour la simplification de textes médicaux.

ABSTRACT. The purpose of automatic simplification is to create version of texts which is easier to understand for a given targeted population. We aim at simplifying medical texts. Usually, lexicon and rules required for the simplification are acquired from parallel corpora. Since such corpora are not available or are very few for French, we propose methods for their creation from comparable corpora. Our method relies on filtering step, which purpose is to keep the best sentence candidates for alignment, and alignment step considered as categorization problem. The aim is to decide whether a pair of sentences is alignable or not. We exploit different types of features (mainly issued from lexicon and corpora). The results obtained allow to build a parallel corpus for biomedical text simplification.

MOTS-CLÉS : Simplification automatique₁, textes médicaux₂, corpus de phrases parallèles₃, constitution de ressources₄.

KEYWORDS: Automatic simplification₁, Medical texts₂, Corpus with parallel sentences₃, Resource building₄.

La simplification automatique a pour objectif de transformer des textes difficiles afin de les rendre plus faciles à comprendre pour un public donné. Nous nous intéressons à la simplification de textes médicaux en français. Les méthodes de simplification automatique procèdent à l'acquisition de lexique et de règles de simplification à partir de corpus parallèles (Shardlow, 2014 ; Saggion, 2017). Comme de tels corpus n'existent pas pour le français ni pour le domaine médical, nous nous proposons de remédier à ce manque. Pour cela, nous proposons une méthode d'alignement automatique de phrases à partir de corpus comparables.

Nous travaillons à partir d'un échantillon provenant d'un corpus comparable du domaine médical en français (Grabar, Cardon, 2018). Nous annotons manuellement une sélection aléatoire de couples de documents pour identifier les cas d'équivalence (les couples de phrases qui expriment le même sens) et les cas d'inclusion (les couples de phrases où l'une phrase inclut le sens de l'autre avec en plus de l'information qui lui est propre). L'accord inter-annotateurs, calculé avec le κ de Cohen (Cohen, 1960), est de 0,76, ce qui correspond à un accord haut. Parmi les 266 couples de phrases alignées retenues, nous obtenons 130 cas d'équivalence et 130 cas d'inclusion.

Pour automatiser cette tâche d'alignement de phrases, nous la modélisons comme un problème de classification binaire: il faut décider si deux phrases sont alignables ou non. Deux obstacles importants se posent: (i) le faible volume de données dont nous disposons ne permet pas d'utiliser les méthodes d'apprentissage profond et (ii) le déséquilibre des phrases alignables et non alignables est extrêmement important.

La méthode d'alignement que nous proposons comporte deux étapes :

1. L'étape de *filtrage* est destinée à éliminer au maximum les phrases non alignables et de réduire le déséquilibre. Pour cela, nous exploitons une méthode inspirée d'un travail existant (Duran *et al.*, 2014). Cette méthode repose sur la comparaison d'arbres syntaxiques à partir de l'analyse en constituants de deux phrases. L'analyse en constituants est effectuée à l'aide de l'analyseur *Benepar* (Kitaev, Klein, 2018). Cette étape de filtrage permet de réduire le déséquilibre des classes de nos données de 8 000 : 1 à 200 : 1, tout en optimisant les phrases alignables.

2. L'étape d'*alignement* consiste à entraîner un classifieur binaire sur les données de référence et à l'appliquer aux données issues du filtrage. Nous exploitons différents types de descripteurs. La plupart de ces descripteurs sont inspirés de travaux sur l'alignement de phrases à partir de corpus comparables (Barzilay, Elhadad, 2003) ou sont des mesures de distance classiques comme la similarité cosinus, les coefficients Dice et Jaccard ou encore la distance d'édition de Levenshtein (Levenshtein, 1966). Nous utilisons également deux descripteurs à base de plongements lexicaux proposés spécifiquement pour l'alignement de phrases en vue de la simplification (Štajner *et al.*, 2018). Nous décrirons plusieurs expériences (le choix de l'algorithme, l'effet du déséquilibre à l'entraînement...) et présenterons les résultats obtenus.

Nous appliquons le meilleur modèle sur des données non vues pendant la phase d'entraînement. Les résultats montrent que la méthode que nous proposons permet d'obtenir un corpus parallèle avec un effort raisonnable de validation manuelle.

Bibliographie

- Barzilay R., Elhadad N. (2003). Sentence alignment for monolingual comparable corpora. In ACL (Ed.), *Proceedings of the 2003 conference on empirical methods in natural language processing*, p. 25–32. Sapporo, Japan. <https://www.aclweb.org/anthology/W03-1004>
- Cohen J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, vol. 20, n° 1, p. 37.
- Duran K., Rodriguez J., Bravo M. (2014, 09). Similarity of sentences through comparison of syntactic trees with pairs of similar words. In *11th international conference on electrical engineering, computing science and automatic control (cce)*, p. 1-6. Campeche.
- Grabar N., Cardon R. (2018). CLEAR – Simple Corpus for Medical French. In *Workshop on Automatic Text Adaption (ATA)*, p. 1-11. Tilburg, Netherlands.
- Kitaev N., Klein D. (2018, July). Constituency parsing with a self-attentive encoder. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia, Association for Computational Linguistics.
- Levenshtein V. I. (1966, February). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, vol. 10, p. 707.
- Saggion H. (2017). *Automatic text simplification* (vol. 32). University of Toronto, Morgan & Claypool.
- Shardlow M. (2014). A survey of automated text simplification. *Int J Advanced Computer Science and Applications*, vol. 1, p. 1-13.
- Štajner S., Franco-Salvador M., Rosso P., Ponzetto S. P. (2018). CATS: A tool for customized alignment of text simplification corpora. In *Lrec*, p. 3895-3903.

Pour une démarche de communication multimodale de données scientifiques : de la recherche documentaire à l'infographie via le *mind mapping*

Sílvia Araújo¹, Radia Hannachi²

1. Centre d'Études Humaines (CEHUM), Université du Minho
Rua da Universidade, 4710-057 Braga, Portugal
saraujo@ilch.uminho.pt

2. Laboratoire HCTI (EA4249), Université de Bretagne Sud
4, rue Jean Zay BP 92116, 56321 Lorient Cedex, France
Radia.Hannachi@univ-ubs.fr

RÉSUMÉ. Dans le cadre de cet article, nous tenons à présenter les résultats d'une expérience pédagogique menée dans un contexte universitaire. Le but de cette expérience est d'introduire les étudiants aux méthodologies actives par le biais d'une démarche pédagogique en trois étapes: i) recherche documentaire de cinq articles scientifiques portant sur une méthodologie active au choix (pédagogie inversée, apprentissage par projet ou par le jeu, entre autres); ii) restitution de la recherche documentaire sous la forme d'une carte mentale; iii) création d'une version infographique de la carte mentale qui présente les atouts et les limites de la méthodologie active traitée. Le processus se met en place à travers un padlet utilisé comme un espace de travail collaboratif contribuant à la mise en œuvre d'un portfolio numérique qui réunit le matériel produit dans le cadre des étapes précédentes. Afin de mettre au jour les perceptions des étudiants face à la démarche proposée, un questionnaire a été administré. Le traitement des données a été effectué suivant une méthode mixte (questionnaire à questions fermées et ouvertes). Les résultats semblent démontrer l'impact positif de la démarche aussi bien au niveau de l'appropriation de concepts didactiques qu'au niveau du développement des littératies numériques. En mobilisant différents outils numériques (mur collaboratif virtuel, carte mentale, création d'une infographie, entre autres), les étudiants s'initient à de nouveaux environnements d'apprentissage qui impliquent un dialogue productif entre recherche documentaire et communication multimodale de cette recherche à l'ère du numérique. Profiter des avancées technologiques en matière de systèmes de l'information pour automatiser la démarche proposée peut paraître tentant à plusieurs égards, mais du point de vue pédagogique, une telle automatisation nous semble limiter le rôle actif des apprenants dans la mesure où elle se concentre uniquement sur le produit final et non sur les stratégies (méta)cognitives à mettre en œuvre pour l'atteindre.

ABSTRACT. In this article, we present the results of a pedagogical experiment conducted in a university context. The goal of this experiment is to introduce students to active methodologies through a pedagogical approach in three stages: i) documentary research of five scientific articles relating to an active methodology chosen by the students (flipped learning, project-based learning, among others); ii) restitution of the documentary research in the form of a mind map; iii) creation of an infographic version of the mind map which presents the strengths and limitations of the chosen methodology. The process takes place through a padlet used as a collaborative workspace, contributing to the implementation of a digital portfolio that brings together the material produced in the previous steps. In order to bring to light the students' perceptions of the proposed approach, a questionnaire was submitted. Data processing was carried out using a mixed-methods approach (closed-ended and open-ended questionnaire). The results seem to show the positive impact of the approach both in terms of the appropriation of didactic concepts and in the development of digital literacies. By using different digital tools (virtual collaborative wall, mind map, creation of an infographic, etc.), the students are introduced to new learning environments that involve a productive dialogue between documentary research and multimodal communication of this research in the digital age. Taking advantage of technological advances in information systems to automate the proposed approach may seem tempting in several respects, but from a pedagogical point of view, such automation can reduce the active role of learners insofar as it focuses only on the final product and not on the (meta)cognitive strategies to be implemented to achieve it.

Mots-clés : méthodologie active, carte mentale, infographie, pensée visuelle, communication multimodale

KEYWORDS: active methodology, mind map, infographic, visual thinking, multimodal communication

1. Introduction

La démarche pédagogique dont nous tenons à rendre compte a été menée dans le cadre d'un cours de didactique du FLE à l'Université de Bretagne Sud avec huit étudiants natifs du français de licence 2 anglais, espagnol, lettres modernes et histoire ayant suivi toute leur scolarité en France. Ce cours a pour objectif de sensibiliser les étudiants à des notions de didactique utilisées dans l'enseignement des langues. Pour les amener à explorer de manière autonome ces notions sans passer par un enseignement magistral, nous avons mis en place un scénario pédagogique qui se focalise sur une pratique des différents processus en jeu dans la recherche, la sélection, l'organisation et la communication de l'information scientifique.

2. Pour une introduction aux techniques de pensée visuelle

Le numérique est porteur de nouveaux modes d'accès, de plus en plus visuels, à l'information : navigation au lieu de lecture linéaire, usage de plus en plus fréquent de supports visuels. Pour répondre aux besoins croissants de synthèse et de simplicité de l'information, les ouvrages traitant des techniques dites de pensée visuelle (*mind mapping, sketchnoting, whiteboarding, scribing, live-sketching*, facilitation graphique, entre autres) foisonnent sur le web (Akoun. 2017 ; Appert, 2018 ; Brand, 2020, entre autres). En associant mots et images, ces différentes techniques permettent

de communiquer de façon visuelle et synthétique un contenu ciblé (Imbeau, 2020 ; Schafer, 2020).

3. Démarche pédagogique centrée sur la création de visuels

Nous décrivons ci-dessous les étapes de mise en place de notre démarche d'introduction à un domaine de spécialité qui implique, tout d'abord, une sélection efficace de l'information au terme d'une recherche documentaire.

3.1. Recherche documentaire

La première étape vise à trouver de l'information sur un sujet donné en procédant de manière exploratoire. Dans un contexte de surabondance informationnelle, ce processus de recherche et de validation de l'information requiert la mise en place d'une méthodologie efficace pour exploiter la richesse documentaire du web scientifique libre (Pochet, 2015). Il existe, aujourd'hui, des millions d'articles scientifiques, de thèses et de mémoires accessibles en ligne. Dès lors, il est important d'amener les étudiants à poser efficacement le sujet à traiter, à sélectionner les concepts/mots clés nécessaires à l'interrogation des sources documentaires et à évaluer la qualité et la pertinence des sources sur la base de critères objectifs. Les étudiants issus de l'enseignement secondaire et qui entrent dans nos universités ne disposent pas nécessairement de compétences solides en matière de littératie de l'information. Pour les aider à entreprendre une recherche d'informations sur le web, celle-ci peut se structurer autour d'une méthode comme le QQQQCP (Qui ? Quoi ? Où ? Quand ? Comment ? Pourquoi ?) qui "est un véritable couteau suisse pour formuler un problème" (Delengaigne et al., 2011 : 35).

3.2. Élaboration de la carte mentale

Pour répondre à cet ensemble de questions simples qui permet d'attaquer un sujet sous différents angles, nous proposons l'élaboration d'une carte mentale qui "stimule en même temps les deux hémisphères et apporte ainsi une véritable mobilité intellectuelle dans le traitement des données" (Delengaigne *et al.*, 2011 : 36). La carte mentale est une hiérarchie de liens entre des données suivant une arborescence dont l'objectif est de structurer et/ou de faire émerger de l'information (Régnard, 2010 ; Mongin, de Broeck, 2019). Cette technique de *mind mapping*, popularisée par le psychologue anglais Tony Buzan dans les années 1970 (Buzan et Buzan, 2003), permet de sélectionner et de hiérarchiser les éléments clés et signifiants provenant de la recherche documentaire menée à l'étape précédente. Lorsqu'il y a beaucoup d'informations à faire passer dans un même espace, le *mind mapping* requiert un esprit de synthèse et un sens de l'organisation des éléments textuels et visuels (Portelli, 2018). En ce sens, la carte mentale, qu'elle soit réalisée en format électronique ou sur papier, peut servir de prototype pour l'infographie.

3.3. *Élaboration de l'infographie*

Avec la carte mentale, il s'agit bel et bien d'organiser et de donner du sens à l'information qu'il convient de considérer. Dès lors, le *mind mapping* permet de réaliser en amont une grande partie du travail de conception de l'infographie. En prônant une approche multimodale qui associe astucieusement mots-clés, pictogrammes et couleurs (Lebrun et Lacelle, 2012), l'infographie constitue, sans aucun doute, un bon point d'entrée pour inciter le public à rechercher davantage d'informations sur un sujet donné. L'infographie "a pour objectif l'organisation rigoureuse et simplifiée de données à destination d'un public avide de fluidité et de clarté" (Canva, s.d). La réalisation de ces visuels nécessite d'identifier les modes de simplification et les critères de choix à la fois sémantiques et esthétiques. Renforcée par le développement du *Big data* et de l'*Open data* (Desfriches Doria, 2015), la visualisation de données constitue, comme le remarquent à juste titre Lehmans & Cardoso (2017), l'un des enjeux importants dans le développement des usages éducatifs des données ouvertes.

La figure qui suit donne un aperçu général des trois étapes dont il vient d'être question :

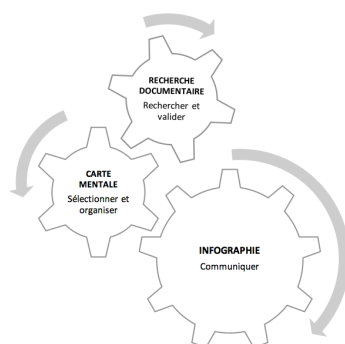


Figure 1. *Vision d'ensemble de la démarche pédagogique*

4. **Mise en place de la démarche dans le cadre d'un cours de didactique du FLE**

Afin d'introduire les quatre binômes d'étudiants de didactique de FLE aux méthodologies actives appliquées à l'enseignement des langues, nous les avons encouragés à suivre la démarche décrite ci-dessus.

Étape 1. De la recherche documentaire ...

En faisant appel à des plateformes et des revues spécialisées dans le champ de la didactique des langues (archive ouverte HAL, lidil, AILE, Alsic, entre autres), chaque binôme a, tout d'abord, été convié à mener une recherche documentaire en lien avec

l'une des méthodologies actives suivantes : binôme 1 : *flipped learning* ; binôme 2 : *project-based learning* ; binôme 3 : *team-based learning* ; binôme 4 : *gamification*.

Étape 2. ... à l'élaboration de la carte mentale ...

Après un cours de présentation des pistes à suivre pour élaborer une carte mentale (Araújo, & Correia, 2020), chaque binôme a été amené à restituer la recherche documentaire menée précédemment sous la forme d'une carte mentale. Pour ce faire, différents outils ont été mobilisés (*GitMind*¹, *Miro*², *MindMeister*³ et *MindMup*⁴) L'utilisation de ces outils n'a soulevé aucun problème majeur. En revanche, deux des étudiants affirment avoir ressenti des difficultés d'ordre méthodologique lors de l'application de la technique de *mind mapping*. Comme le souligne, à ce propos, l'un d'entre eux, il n'est pas simple « de convertir toutes les informations retenues lors de la recherche documentaire en parties claires et concises pour la carte mentale » (Étudiant 6, ci-après E6).

Étape 3. ... pour parvenir à l'infographie

À l'aide d'un des outils proposés (*Canva*⁵, *Visme*⁶, *piktochart*⁷, entre autres), chaque binôme a finalement procédé à la création de la version infographique de la carte mentale réalisée à l'étape précédente. Au cours de cette dernière étape, les étudiants affirment avoir été confrontés à des défis d'ordre technique (cf. *infra* E5, E6a), méthodologique (cf. *infra* E3) et/ou purement esthétique (cf. *infra* E4, E6b) :

E5 : *Nous n'avions jamais réalisé d'infographies auparavant et c'était, pour nous, la première fois que nous utilisions le site Canva. Cela a donc pris du temps de prendre en main le logiciel. [...] Au fur et à mesure que nous manipulions le site internet on commençait à s'y habituer et il était donc plus facile de faire des ajustements.*

E6a : *Il fallait également réussir à s'adapter au logiciel de création de cette infographie que l'on n'avait jamais utilisé auparavant.*

E3 : *Je n'avais jamais pratiqué d'infographie, alors j'ai dû apprendre doucement comment elle marchait, c'est sûrement la seule difficulté que j'ai rencontrée.*

E4 : *Et le choix des couleurs pour que le texte soit lisible et que les couleurs s'accordent bien. Nous avons donc essayé plusieurs combinaisons.*

¹ <https://gitmind.com/>

² <https://miro.com/>

³ <https://www.mindmeister.com/pt>

⁴ <https://www.mindmup.com/>

⁵ <https://www.canva.com/>

⁶ <https://www.visme.co/>

⁷ <https://piktochart.com/>

E6b : *La principale difficulté rencontrée lors de l'élaboration de notre infographie a été de réussir à mettre en place quelque chose qui attire l'œil de ceux qui la regardent et qui fasse en sorte qu'ils aient envie de la lire. [...] Pour ce faire, on a surtout demandé conseil autour de nous pour savoir si le visuel était correct et s'il attirait le regard et l'attention.*

Sur une échelle de likert de 1 (Très insatisfait(e)) à 5 (Très satisfait(e)), chacun de ces binômes a procédé à l'autoévaluation de l'infographie réalisée (étape 3) à partir de la carte mentale (étape 2) selon sept paramètres : P1: Qualité de l'information (les informations essentielles sont présentes et le sens de lecture est pertinent); P2: Qualité du langage utilisé (langage clair, en rapport avec le sujet, sans coquilles/erreurs grammaticales); P3: Impact visuel agréable; P4: Titres et entêtes clairs et mis en évidence; P5: Textes/images (complémentaires); P6: Mise en page aérée; P7: Lisibilité de l'infographie (choix de la typographie, couleurs contrastées pour le fond et les textes, etc.). Il est à noter que tous les paramètres ont obtenu une moyenne arithmétique située entre 4,00 et 4,29, à l'exception du paramètre 5 qui a obtenu une moyenne (légèrement inférieure) de 3,57. Dans l'ensemble, les étudiants ont donc évalué très favorablement leur production finale, malgré les difficultés indiquées plus haut. Il y aurait ici tout un champ de recherche intéressant en ce qui concerne les paramètres à prendre en compte pour réaliser une infographie efficace. À en juger par les commentaires tissés par les étudiants pour choisir l'infographie la plus réussie au sein des binômes, il semblerait que les paramètres P3 (impact des couleurs) et P6 (quantité d'information visible) mentionnés ci-dessus soient déterminants dans le choix.

5. Exemplification de la démarche à travers quelques productions

Tous les binômes affirment avoir transféré, sur leur infographie, la plus grande partie des informations figurant sur leur carte mentale, sans que l'ordre de présentation de ces informations soit nécessairement respecté d'un visuel à l'autre.

5.1. Méthodologie active – Flipped Classroom

Nous illustrons ci-dessous le passage de la carte mentale (cf., *infra*, Fig. 2) à l'infographie (cf., *infra*, Fig. 3) proposé par le binôme ayant travaillé sur la classe inversée à partir d'une recherche documentaire axée sur cinq articles scientifiques. Comme on peut le constater, outre l'introduction et la conclusion, la carte mentale présente quatre sections (reliées au nœud central par une branche) qui mettent en avant les potentialités (section 1), les conditions de mise en place de la méthodologie (section 2), ainsi que les outils (section 3) et les limites (pédagogiques et/ou logistiques) qui peuvent être rattachés au *flipped classroom* :

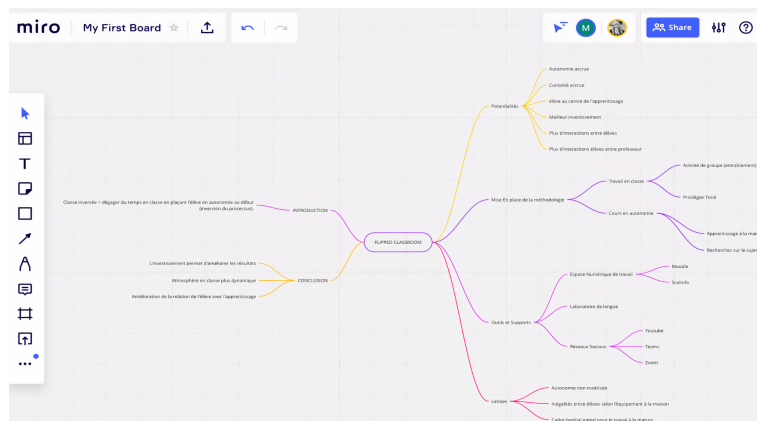


Figure 2. Carte mentale élaborée à partir d'une recherche documentaire portant sur la classe inversée

Élaborée à partir de la carte mentale précédente, l'infographie qui suit reprend grosso modo les quatre sections dont il vient d'être question, mais opte pour un ordre de présentation légèrement différent : la section 2 (Mise en place de la méthodologie) apparaît avant la section 1 (Potentialités/bénéfices) et la section 4 (Limites) avant la section 3 (Outils) :



Figure 3. Infographie élaborée à partir de la carte mentale portant sur la classe inversée

5.2. Méthodologie active - Team-based learning

À l’instar de l’exemple précédent, nous présentons la carte mentale (cf., *infra*, Fig. 4) élaborée par le binôme du *team-based learning* (TBL), puis la version infographique (cf., *infra*, Fig. 5) de cette carte mentale qui s’efforce de condenser les principales informations recueillies lors de la recherche documentaire :

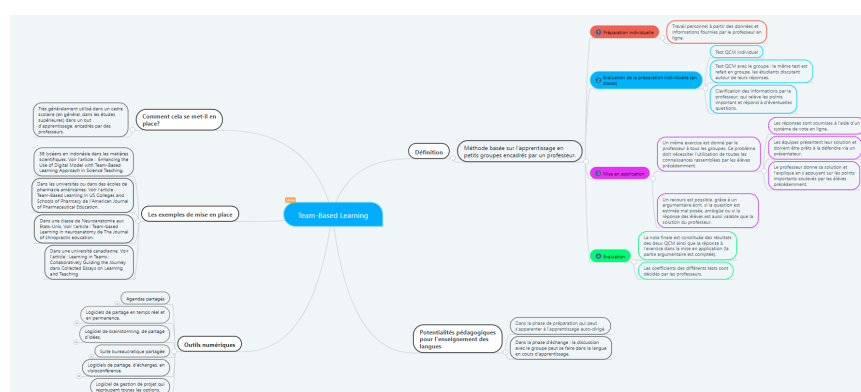


Figure 4. Carte mentale élaborée par le binôme du *team-based learning*

Il est à noter que l'infographie qui suit reprend fondamentalement deux des sections représentées sur la carte mentale qui précède, à savoir la section rattachée à la définition du TBL ainsi que la section liée aux outils numériques nécessaires à la mise en place de la méthodologie :

Figure 5. Infographie faisant écho à la carte mentale portant sur le *team-based learning*

6. Conclusion

L'expérience pédagogique dont il est question ci-dessus porte sur un échantillon fort réduit. Précisons, néanmoins, que d'autres expériences sont en cours pour approfondir la réflexion ébauchée ici. À partir des résultats préliminaires présentés ci-dessus, il nous semble essentiel de mettre en avant l'importance de la technique de *mind mapping* comme un préalable indispensable au développement des compétences en littératie universitaire. Comme nous avons pu en rendre en compte dans les pages qui précèdent, l'élaboration de cartes mentales en articulation avec une méthodologie de recherche documentaire s'inscrit dans une démarche qui amène les étudiants à être actifs sur plus d'un plan : il s'agit bel et bien de lire et de comprendre des textes (scientifiques) pour en extraire les informations importantes et les organiser sous un nouveau format de communication visuelle à l'aide d'outils numériques disponibles en ligne. Appliquée à l'éducation, cette démarche de simplification et vulgarisation scientifique basée sur la création de contenus multimodaux (tels que la carte mentale ou l'infographie, par exemple) vise donc à faciliter aussi bien le développement de la capacité d'analyse et de synthèse que celui de compétences transversales telles que l'autonomie, la créativité, la collaboration et la littératie numérique. Le développement de ces compétences nous semble résider dans la mise en place d'une méthodologie qui donne un rôle actif aux apprenants. On peut estimer néanmoins les avantages que pourrait avoir une automatisation du passage de la recherche documentaire à la carte mentale, voire de cette dernière à l'infographie dans le cadre de la simplification et vulgarisation scientifique. Cette quête d'automatisation recouvre des défis importants pour des domaines divers tels que la recherche d'information et le traitement automatique des langues, entre autres. Il s'agit non seulement de mettre en lumière les aspects à considérer pour concevoir un workflow de *mind mapping* automatique (Chen, Krishnamurthy, 2020 ; Chen *et al.*, 2020) mais de réfléchir également à des pistes d'intégration pédagogique d'outils de génération automatique de visuels (cartes mentales ou autres) qui puissent résulter des avancées en matière de système d'information.

Remerciements

Cet article a été partiellement réalisé dans le cadre du projet PortLinguE (PTDC/LLT-LIG/31113/2017), financé par le Fonds Européen de Développement Régional (FEDER) et par la Fondation pour la Science et la Technologie (FCT) du Portugal.

Bibliographie

- Akoun A. (2017). *Travailler avec le sketchnoting comment gagner en efficacité et en sérénité grâce à la pensée visuelle*. Paris, Eyrolles.
- Appert É. (2018). *Penser, dessiner, révéler : toutes les méthodes pour accompagner les idées, les équipes et la vie par le dessin*. Paris, Eyrolles.

- Araújo S., Correia A. (2020). Using speechmaking and consecutive interpreting as tools to help students develop writing and public speaking skills: a hybrid teaching methodology based on mind mapping. *inTRAlinea Special Issue: Technology in Interpreter Education and Practice*. Stable URL: <http://www.intralinea.org/specials/article/2515>
- Brand W. (2020). *Visual thinking : la méthode qui révolutionne vos idées sketchnoting, scribing, facilitation graphique pour tous*. Paris, Éditions Vuibert.
- Buzan T., Buzan B. (2003). *Mind Map - Dessine-moi l'intelligence*. Paris, Les Éditions d'Organisation, Collection Les guides Buzan.
- Canva (s.d). *Les 9 clés d'une infographie réussie*. https://www.canva.com/fr_fr/decouvrir/infographie-reussie/
- Chatenet L., Cardoso S. (2020). Du graphique à l'infographie. De l'art de faire parler les images. *Interfaces numériques*, 9 (3). <http://dx.doi.org/10.25965/interfaces-numeriques.4408>
- Chen, T. J., Krishnamurthy V. R. (2020). Investigating a Mixed-Initiative Workflow for Digital Mind-Mapping. *Journal of Mechanical Design*, 142(10): 01404. doi: 10.1115/1.4046808
- Chen T. J., Subramanian S. G., Krishnamurthy V. R. (2020). QCue: Queries and Cues for Computer Facilitated Mind-Mapping. *In Proceedings – Graphics Interface*, pp. 125 - 136. University of Toronto, 28 - 29 May 2020, Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine. <https://doi.org/10.20380/GI2020.14>
- Delengaigne X., Deschamps C., Mongin P. (2011). *Organisez vos données personnelles : L'essentiel du Personal Knowledge Management*. Eyrolles Éditions d'Organisation.
- Desfriches Doria, O. (2015) Quels dispositifs numériques pour appréhender la datavisualisation ?. *I2D – Information, données & documents, A.D.B.S., 2015, Dossier Datavisualisation, des données à la connaissance*, 52 (2), p.88. (10.3917/i2d.152.0054). (halshs-01701145)
- Imbeau L. (2020). *Comprendre et communiquer la science*. 3e édition. Presses de l'Université Laval.
- Lebrun M., Lacelle N. (2012). Le Document Multimodal : Le Comprendre et le produire en classe de français. *Repères*, n° 45, p. 81-95.
- Lehmans A., Cardoso S. (2017). Datavisualisation des données ouvertes et design pédagogique dans les formats de connaissances. *Conférence internationale H2PTM'2017 : Le numérique à l'ère des designs (contenus, interactions, espaces, environnements, services, objets, œuvres, programmes.) : de l'hypertexte à l'hyper-expérience*, Oct 2017, Valenciennes, France. (hal-01638054)

- Mongin P., de Broeck, F. (2019). *Enseigner autrement avec le Mind Mapping. Cartes mentales et conceptuelles* - 2e édition. Collection : La Boîte à Outils du professeur. Dunod.
- Pochet B. (2015). *Comprendre et maîtriser la littérature scientifique*. Presses agronomiques de Gembloux.
- Portelli M. (2018). *Les cartes heuristiques : un outil d'individualisation de la synthèse*. Education. dumas-01831807f
- Régnard D. (2010) Apports pédagogiques de l'utilisation de la carte heuristique en classe. *Études de Linguistique Appliquée*, n° 2, p. 215-22.
- Schafer V. (2020). *Information et communication scientifiques à l'heure du numérique*. Paris, CNRS Éditions.

Controlled language for text simplification

Concepts and implementation

Mike Unwalla¹

1. TechScribe

52 Stanwood Crescent, Sheffield, S6 5JB, UK

mike@techscribe.co.uk

RÉSUMÉ. Dans le commerce et l'industrie, de nombreuses organisations utilisent des langues simplifiées, par exemple "plain English" et "lenguaje claro" [espagnol]. Pour la documentation technique, critique pour la sécurité, "plain language" n'est pas toujours suffisant et certaines organisations utilisent une langue contrôlée. L'ASD-STE100 Simplified Technical English est la spécification d'une langue contrôlée. Dans cet article, nous présentons TechScribe, un logiciel conçu pour vérifier automatiquement la conformité d'un document aux règles ASD-STE100. De nombreuses règles de l'ASD-STE100 sont applicables à la simplification des textes scientifiques. Pour le démontrer, dans la mesure du possible, cet article est conforme à la norme ASD-STE100.

ABSTRACT. In commerce and industry, many organizations use plain language, for example, 'plain English' and 'lenguaje claro' [Spanish]. For safety-critical documentation, plain language is not always sufficient, and some organizations use controlled language. ASD-STE100 Simplified Technical English is a specification for a controlled language. In this paper, we present the TechScribe term checker for ASD-STE100, which checks a document for conformity to ASD-STE100. Many of the ASD-STE100 rules are applicable to the simplification of scientific texts. To show that, this paper conforms to ASD-STE100 as much as possible.

MOTS CLÉS: ASD-STE100 Simplified Technical English, CL, CNL, langage (naturel) contrôlé, ambiguïté grammaticale, spécification, simplification de texte, terminologie, validation par logiciel.

KEYWORDS: ASD-STE100 Simplified Technical English, CL, CNL, controlled (natural) language, grammatical ambiguity, specification, text simplification, terminology, validation with software

1. Introduction: commercial and industrial text simplification methods

English is an important language for international communication, but like other natural languages, English *can be* complex, unclear, and confusing. In business and commerce, ‘plain language’ and ‘plain English’ are important (PLAIN, 2021). Many organizations use plain language. Plain language is good, but for some types of document, controlled language (CL) is better. Plain language is a set of guidelines, not a specification (Unwalla and Dodd, 2017).

CL is a “*subset of natural languages whose grammars and dictionaries have been restricted in order to reduce or eliminate both ambiguity and complexity*” (ISO, 2015). For English, more than 100 CLs are available (Kuhn, 2013).

In the technical documentation industry, a well-known CL is ASD-STE100 Simplified Technical English (ASD, 2017). It is an international specification for the preparation of technical documentation in a controlled language. But, CL is also applicable to other genres, such as Bible texts (Betts, 2005).

Research at SAS resulted in Global English (Kohl, 2008). Kohl writes that Global English is an alternative to translation if the readers are ‘reasonably proficient’ in English. Global English identifies the grammatical structures that cause problems for non-native readers. Global English is more controlled than plain English, but not as controlled as a CL.

This article introduces the TechScribe term checker for ASD-STE100 Simplified Technical English (www.simplified-english.co.uk), which helps writers to make sure that a document conforms to ASD-STE100. Rules from ASD-STE100 are applicable to the simplification of scientific texts.

To show that simple text is applicable to technical subjects, this paper conforms to ASD-STE100 as much as possible.

2. ASD-STE100 Simplified Technical English

English has a rich vocabulary and that richness makes it good for literature and poetry. But, that richness causes problems with technical texts:

- phrasal verbs cause problems for non-native readers (Thrush, 2001);
- synonyms can confuse readers;
- one word (sequence of characters) can have different parts of speech, which means that one sentence can have more than one interpretation;
- one word can have different meanings. The verb ‘fall’ can mean ‘move down because of the effect of gravity’ or ‘decrease’;

The ASD-STE100 rules help writers to make the text clear and consistent.

ASD-STE100 has two parts: (1) a dictionary of approved and unapproved terms and (2) a set of writing rules.

The approved words in the dictionary are basic words that are applicable to all technical texts. Many unapproved words are synonyms of approved words. For example, the adjective ‘big’ is not approved. Writers must use ‘large’. But, if your text is about ‘big data’, then ‘big data’ is a technical term. The sentence “*Hadoop is used to analyse big data sets*” has only one meaning. (‘Big data sets’ means ‘sets of big data’, not ‘data sets that are large’.)

The writing rules help writers to make text as clear as possible. Some examples are in section 5 of this article. ASD-STE100 has more than 350 pages of rules. Software can help writers to comply with the specification.

ASD-STE100 tells writers to use terminology consistently, but it does not tell them how to get the terminology. Terminology mining and term selection are necessary to use ASD-STE100 fully.

3. Checker for ASD-STE100

In the early 2000s, commercial checkers for ASD-STE00 (AECMA SE at the time) were expensive. TechScribe started to customize LanguageTool (www.languagetool.org) to develop a low-cost checker for ASD-STE100.

The checker has two parts: (1) a set of rules that make a disambiguator, which identifies the part of speech that a word has and (2) a set of rules that identify text that does not conform to ASD-STE100.

The rules are written in XML. The design of the checker is on www.simplified-english.co.uk/design.html.

Unlike in LanguageTool, the disambiguator adds readings. It asserts that a word has a part of speech, for example, IS_NOUN. Usually, if a word has two different parts of speech, for example, IS_NOUN and IS_VERB, then the disambiguation is not correct. Rules find words that have two parts of speech, and thus the disambiguation errors can be corrected.

Some limits of the checker

The disambiguator uses patterns in text (equivalent to n-grams), not full sentence parsing. Thus, it does not always correctly identify the part of speech.

The source text must be correct English. If source text is not correct, then disambiguation is not correct. Thus, part-of-speech rules will not give a correct analysis.

The checker can analyse only a subset of the ASD-STE100 rules. For example, rule 6.5 is ‘make sure that each paragraph has only one topic’. The checker has no semantic knowledge. It analyses only grammar.

4. Some sentences are grammatically ambiguous

For some sentences, without contextual or semantic information, we cannot know the correct analysis. In the sentence “*This lets you adapt it easily to weather changes*”, we can parse ‘weather’ as a verb (with the meaning of ‘to overcome’) or as a noun in a noun cluster.

The checker has no semantic knowledge. It does not know that ‘drunk’ has a different part of speech in these sentences:

- The water was drunk. (‘Drunk’ is the past participle in the passive voice.)
- The waiter was drunk. (‘Drunk’ is an adjective that describes the state of the waiter.)

More examples are on www.simplified-english.co.uk/analysis.html.

5. Controlled language for scientific texts

To use a controlled language fully is not easy. Terminology mining and term selection are necessary to make sure that only approved terms are used. But, many rules of ASD-STE100 are applicable to all types of descriptive texts. You can use these rules without the terminology rules. Examples:

- rule 1.2, use approved words only as the part of speech given;
- rule 1.3, use approved words only with their approved meanings;
- rule 1.11, do not use different technical names for the same thing;
- rule 3.5, do not use the ~ing form of a verb (present participle);
- rule 9.3, when you use two words together, do not make phrasal verbs;

“*Complex concepts don't mandate complex explanations. Tangled prose reflects a tangled writer with a loose grip on their subject matter*” (Wardak 2020).

References

- ASD (2017). *ASD-STE100 Simplified Technical English: International specification for the preparation of technical documentation in a controlled language*, ASD, Brussels.
- Betts R. (2005). Wycliffe Associates' EasyEnglish. *Communicator*, Spring 2005, p. 28–31.
- ISO (2015). *Language resource management — controlled natural language (CNL) — Part 1: Basic concepts and principles (ISO/TS 24620-1)*, ISO, Geneva.
- Kohl J. R. (2008). *The Global English Style Guide: Writing Clear, Translatable Documentation for a Global Market*, Cary, SAS Institute.
- Kuhn T. (2013). *A Survey and Classification of Controlled Natural Languages*. <http://attempto.ifi.uzh.ch/site/pubs/papers/kuhn2013cl.pdf>

- PLAIN (2021). *What is plain language?* <https://plainlanguagenetwork.org/plain-language/what-is-plain-language/>
- Thrush E. A. (2001). Plain English? A Study of Plain English Vocabulary and International Audiences, *Technical Communication*, 48(3), p. 289–296.
- Unwalla and Dodd (2017). The case for ASD-STE100 Simplified Technical English. *Current Practices and Trends in Technical and Professional Communication*. Croydon, ISTC, p. 29–56.
- Wardak T. (2020) Tech-comms as linguistic Tetris, *Communicator*, Winter 2020, p. 30.

Could automatic text simplification assist correction-revision of scientific texts written by non-native English speakers?

Helen McCombie

*Bureau de Traduction de l'Université, Université de Brest
Faculté Segalen, 20 rue Duquesne, 29238 Brest CEDEX, France
helen.mccombie@univ-brest.fr*

RESUME. Je propose ici d'explorer les possibilités offertes par les technologies de simplification automatique de texte dans le contexte de mon travail de correction-révision d'articles scientifiques en anglais. Le besoin constant de définir du vocabulaire nouveau constitue une part importante de la tâche du réviseur confronté à un large spectre de sujets spécialisés. J'examinerai donc les différentes manières d'aborder ce problème. Bien que les outils de simplification de texte destinés aux apprentis lecteurs ou à la création de textes destinés au grand public à partir de textes très techniques ne répondent pas précisément aux besoins d'un réviseur, ils pourraient néanmoins ouvrir la voie à un nouveau workflow. Cet exposé n'est pas une étude expérimentale, mais plutôt un premier pas pour s'interroger sur la manière dont ces outils pourraient faciliter la correction-révision d'articles scientifiques, sur l'aide qu'ils pourraient apporter à un utilisateur humain et sur leur capacité à s'entraîner eux-mêmes.

ABSTRACT. Considering ways to apply automatic text simplification techniques, I present my day-to-day method for correction-revision of scientific articles in English and explore how these tools could help. The constant need for the definition of new vocabulary emerges as important for a reviser faced with correction of texts on a wide range of specialist subjects, so I examine the ways in which this problem can be approached. While tools aimed at learner readers or for creating texts for general readership from highly technical ones are not an exact fit to the needs of a reviser, they could pave the way to a new workflow. This paper is not an experimental study but rather a first step to asking how these tools could assist in a process I use in my work in correction-revision of scientific papers, how they could help to train a human user or, potentially, themselves.

MOTS-CLÉS : correction-révision, simplification du texte, rédaction scientifique, articles scientifiques, enrichissement du vocabulaire, identification de mots complexes

KEYWORDS: correction-revision, text simplification, scientific writing, scientific articles, vocabulary building, complex word identification

1. Introduction

Automatic text simplification applications have been created to help learner readers understand and expand their vocabulary through the text they discover and to make information more accessible to those with cognitive or learning impairments (Caroll et al., 1998; Saggion, 2017). Such programs function by decomplexifying texts on lexical or syntactic bases.

The tasks involved in correcting the written English of scientific articles produced by non-native English speakers before publication require a good level of English on the part of the reviser, but also an understanding of the subject of the text in order to help the author/s convey their intended message. A reviser cannot be a specialist in all fields but is nonetheless frequently faced with highly specialised texts in domains that are new to them. Hence, extensive research is often required in order to revise correctly and in the style of the relevant field. This is an analogous need to the one of the learner reader.

2. Multi-subject correction-revision at a university translation centre

In this paper I will describe my own work process as a reviser of texts at the Bureau de Traduction de l'Université <http://www.univ-brest.fr/btu/> at Brest University (Université de Bretagne Occidentale, UBO) and how I consider this relates to automatic text simplification.

The response of many manuscript correction services to the vast range of subjects in their client's work is to first find a native English speaking reviser with background knowledge or professional experience for the subject of the manuscript to be corrected. They thus work with a large number of service providers, each offering different skills and knowledge. These revisors already have the appropriate lexical and stylistic knowledge of the subject matter they are dealing with and learning time is saved. In other words, they have 'personal dictionaries' of the subjects. For example, on their websites, Wiley Services say they will be '*pairing you [the author] with a native English-speaking editor, with expertise in your subject area*'; Nature employ '*Native English editors with a research or clinical background in your subject areas*'; or, at Enago, '*every paper submitted to us is matched to a highly specialized subject-area expert to ensure technical accuracy.*' It should also be noted that although Enago now offer AI-based correction, they still rely on a broad base of experts. The possible relation of these services I will return to later in section 3.

As a small team (three full-time employees and an MCF manager) the BTU caters for the language service needs of the university and its partners in terms of written translation and interpreting (mostly EN<>FR), correction-revision (of EN),

and associated tasks like subtitling. I specialise in correction-revision of scientific texts, on a broad spectrum of subjects. Although the BTU works with external independent service providers, this is mainly for translation and interpreting, especially outside the main language pair. Little correction-revision is presently out-treated as this is not cost-effective. Learning the vocabulary, characteristics of expression and concepts involved in subjects that are new to me is therefore an integral part of my work.

Scientific texts for revision at the BTU come from all faculties of UBO but also the partners such as the University of Nantes and institutions like INRAE, Ifremer, and CNRS. Subjects span from mathematics to social sciences. Unlike the experts mentioned by the companies above, I frequently come to a text without a great deal of background and therefore repeatedly find myself in the position of an untrained beginner. Figure 1 presents an overview of my workflow.

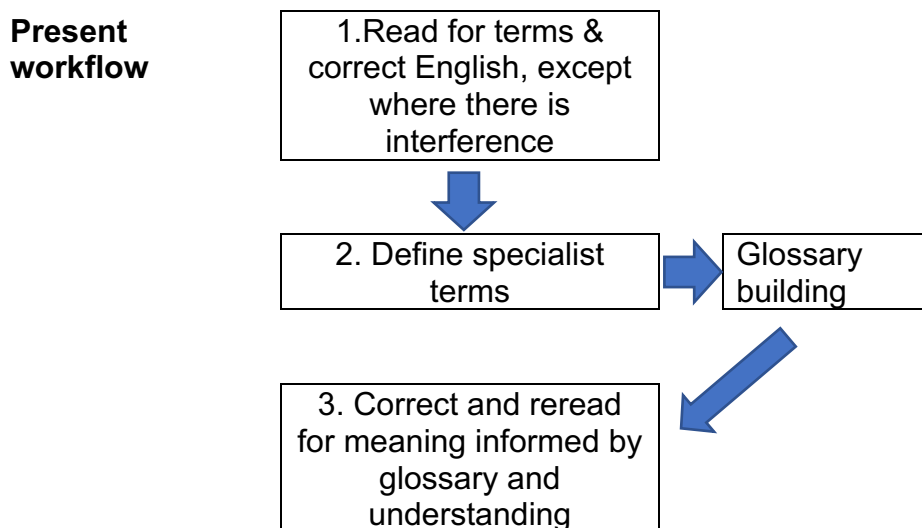


Figure 1. Schema of a sequence of work for correction-revision of a scientific article on an unfamiliar subject

2.1. Stages of correction-revision

When I start reading a paper (Figure 1, step 1), my aim is to test the strength of the article to explain and inform. The things which are simple grammatic mistakes I correct, the terms I do not yet know I mark in one colour, the places where insufficient information is provided to understand the author's intent, regardless of technical terms. I continue to the end of the paper unless these difficulties reach a point that hampers my progress, in other words if I reach a knowledge threshold (O'Reilly *et al.*, 2019). Sometimes, information provided later in a manuscript solves earlier questions. One might argue that the true target audience is probably reading not for overall fluidity but to extract key information as might any engine designed to facilitate this task (e.g. Curiel *et al.*, 2021); therefore, a future objective may be AI-reader optimization. For the moment, however selection of papers for publication remains in human hands (editorial boards and peer reviewers).

In step 2 (*It pays to improve your word power!*¹) I then define terms (often building a glossary of definitions in a spreadsheet (which is useful both for repeated uses within a single work and for future manuscripts by returning customers) based on a number of tools and resources and to integrate these into my understanding. I then make any corresponding corrections (step 3).

The resources I use in step 2 are mainly online scientific journals in the subject area of the text to be corrected via the UBO university library, particularly those cited by the authors, Google Scholar ISTEEX, NCBI, arXiv and Research Gate. I also use search engines, most often to check for frequency of usage (e.g. words that are more or less frequently hyphenated).

After these stages are complete I listen to a text with text to speech software (MacOS Catalina 10.15 tool) for fluidity. Here I recognise that I am comparing with my inner model of what sounds right (no explicit consideration of rules) compared with the English I have already experienced. I am searching here for a sense of normality. This is the most instinctive part of the process, where my action might be comparable with AI style evaluation.

2.2. Similarities between the aims of text simplification tools and my correction process

Text simplification potentially serves not only learners, but a readers wishing to ease this task and obtain information faster (Djamasbi *et al.*, 2016) – especially as smartphones can be less comfortable to read on –, those wishing to get e.g. medical information in lay terms (Kandula *et al.*, 2010) or understand the true scientific

¹ The name of a monthly column in the popular American magazine Reader's Digest whose purpose was vocabulary building. Reader's Digest itself was born from the idea of condensing (manually producing shortened versions) of popular books and articles.

discovery behind a sensationalist news story (Ermakova *et al.*, 2021). None of these reasons is a perfect match to my needs as, for example the replacement of specialist term known to the target readership by a simpler would be unhelpful: my purpose is rather to have a quick method of understanding a subject. A contextual dictionary could therefore be one useful approach.

Proposed solutions include right click access to a search engine. We find this in 'Smart Lookup' of MS Word) or Kindle a dictionary or link to Wikipedia. The former englobes the whole Internet and frequently puts commercial links first the latter works on PDF files (one solution is to run Kindle in parallel on a pdf of the text to be corrected alongside the editable version in the word processor. This is still a term by term approach.

A prepared simplification tool could thus identify and take make propositions in the case of a complex word. My present approach being simply to ask myself if I've met the word before and understand it - whether it exists in my personal 'mental dictionary'. In contrast, the automatic detection of specialist terms is often made by frequency but has become a subject of diverse methods for automatic detection (Sikka and Mago, 2020; Shardlow *et al.*, 2021). Automatic simplification aims to supply either a simpler word or an explanation. The latter tends to lengthen the text as some words have no simpler equivalent, being themselves a shorthand for a larger concept.

3. How can simplification tools help in correction-revision of scientific articles?

One aim of automatic text simplification is to offer definitions of difficult words to enable easier understanding. This form of lexical simplification relies on a defined wordlist or dictionary. It involves the aspect of text simplification that aims to enhance understanding rather than remove complexity. There is, therefore, a parallel between the needs of a learning reader, using such as boosting their vocabulary, and a reviser who needs to grasp the meaning of certain words and turns of phrase rapidly in order to correct and or improve a text in a way compatible with its intended message.

Although the difficulty level of these two types of readers (learners vs revisers) is different, the reviser having already a strong grasp of the language, the need for comprehension of new terms is the same. I want to explore the ways in which a simplification application could be harnessed for use on a variety of specialist fields by substituting the connected dictionary of definitions (e.g. a medical dictionary) and increasing the threshold of difficulty of known/unknown words (O'Reilly *et al.*, 2019) which studies often base on their frequency (or absence) from a general corpus.

There also exists, however, a major difference between the types of texts. Those read by revisers contain errors, their correction being the objective of the reviser's

work. Such errors may exacerbate understanding difficulties by interference. If, for example, there is an unknown term or expression in the text to be revised, whether it consists of one word or several, this may be recognised as an error, by a reviser or by an automatic grammar checker, when it is simply specialist language. Therefore, a part of the reviser's job is to distinguish the words and turns of phrase specific to the subject of the text and unusual enough to be questioned as possible English errors from actual errors that impede comprehension. Furthermore, common words may be used in uncommon ways, e.g. 'head of a muscle', which will not be detected by a grammar- or spell-checker and may still be a misuse. If the reviser does not know the term (in this case is a real one), they have to check it as a possible error, establish it exists in the field. This is all the more the case with multi-word terms or subject-specific terms. All this is part of the glossary-building task.

The reviser needs to put themselves, as best as they can, in the position of the target reader who has a knowledge of the subject, for this, the automatic detection of specialist terms would be more helpful than a standard spelling grammar checking tool. At the same time, the target reader, if the text is a scientific article, will probably have a greater understanding of the subject matter and its specific uses of language than does a reviser.

It is important therefore to be able to quickly gain specific lexical and semantic knowledge of a subject and its writing style in order to distinguish language errors from specialist language as well as to understand the concepts used in the text, which often amount to having a simple definition of the terms and expressions in order to convey the message of the text.

Text simplification tools, intended to improve readability and understanding, presently offer definitions but only simple ones, so trials using freely useable applications such as Rewordify (<https://rewordify.com>) provide something too simple. Similar code to that behind these applications, linked to more sophisticated lexical resources (e.g. a glossary of electronics terms or a medical dictionary), could potentially provide tools that could assist the sometimes complex task that is correction-revision of highly specialised texts.

Alternatively, AI-based tools may be trained on a large general corpus and depending on their agility can then apply their acquired knowledge in diverse contexts, e.g. WebText and GPT-2 (Radford *et al.*, 2019). In a scientific context the use of Google Scholar (Brezina 2012), or COCA-Academic (Davies, 2013) has already been proposed. Recent workshop discussions highlighted (ISTEX Day 19 Jan 2021²) ISTEX to study word uses evolution and annotations and could offer interesting perspectives to train AI tools.

² videos available (last accessed 10 May 2021) at <https://www.youtube.com/channel/UCyow0tVlCRcJwRfk9p7q8Uw>

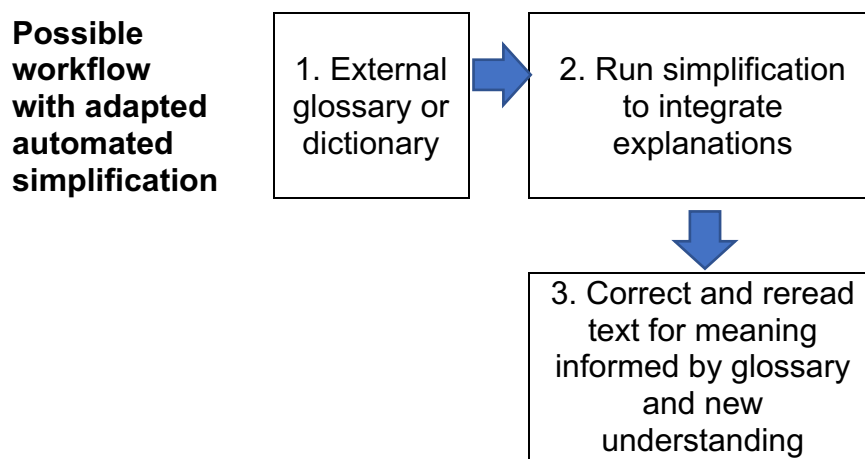


Figure 2. A possible workflow for correction-revision with a simplification tool based on an external dictionary

Considering commercially available solutions, Enago is a large company offering scientific editing services. It also now offers AI-based services 'AuthorOne', including 'Trinka' an academic English checker created by the Crimson AI company. As this is a private company, I do not know anything about the interrelation of these activities, but would think that Enago, as an editing platform has access to a corpus of paired uncorrected and human corrected texts in a wealth of domains. To my mind, this provides a resource of likely errors and possible solutions to learn from could be used to train an AI.

4. Conclusion

I can see from this short exploration of text simplification in relation to my needs as a revisor that there are two major routes I could take. The first is to use a form of contextual dictionary that I harness to my correction environment for when I need to know a word (section 2.2). This would need to be specific and specialised and would, therefore, be difficult to create for the simultaneously broad but specialised demands of my work situation. Alternatively, I could allow an automated text simplification engine to edit the text for me and propose meanings for terminology it considers complex (Figure 2.), treating a text as a whole and followed by post-editing. In either case, simplification technologies could likely contribute to improving my process of manuscript correction-revision.

References

- Brezina, V., (2012). Use of Google Scholar in corpus-driven EAP research. *Journal of English for Academic Purposes* vol 11, n°4, Dec 2012, p 319-331.
- Caroll J., Minnen G., Canning Y., Devlin S., Tait J. (1998) Practical Simplification of English Newspaper Text to Assist Aphasic Readers *Proc. of AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*.
- Curiel A., Gutiérrez-Soto C., Rojano-Cáceresb J.R. (2021). An online multi-source summarization algorithm for text readability in topic-based search. *Computer Speech and Language*, vol 66, Mar 2021, p.101143.
- Davies, M. (2013). Google Scholar and COCA-Academic: Two very different approaches to examining academic English, *Journal of English for Academic Purposes*, Vol 12, n° 3, p. 155-165.
- Djamasbi S., Rochford J., DaBoll-Lavoie A., Greff T., Lally J., McAvoy K. (2016) Text Simplification and User Experience. In: Schmorow D., Fidopiastis C. (eds) *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience*. AC 2016. Lecture Notes in Computer Science, vol 9744. Springer, Cham.
- Ermakova L., Bellot P., Braslavski P., Kamps, J. Mothe J. et al.. Text Simplification for Scientific Information Access: *CLEF 2021 SimpleText Workshop. 43rd edition of the annualBCS-IRSG European Conference on Information Retrieval : Advances in Information Retrieval (ECIR2021)*, Mar 2021, Lucca (virtual), Italy. fihal-03121986f
- Funk P. (1968). *It pays to increase your word power*, Funk and Wagnalls, New York.
- Kandula, S. Curtis, D. Zeng-Treitler, Q., A Semantic and Syntactic Text Simplification Tool for Health Content. *AMIA Annu Symp Proc. 2010*, p.366–370.
- O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. *Psychological Science*, 30(9), 1344-1351.
- Radford A., Wu J., Child R., Luan D., Amodei D., Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
- Saggion H. (2017) Automatic Text Simplification. *Synthesis lectures on human language technologies*, vol. 10 n°1, p. 1-137
- Shardlow M., Evans R., Zampieri M. (2021). *Predicting Lexical Complexity in English Texts*. ArXiv, <https://arxiv.org/pdf/2102.08773.pdf>
- Sikka P., Mago V. (2020). A Survey on Text Simplification. *J. ACM* 37, 4, Article 111 (June 2020), 26 pages. <https://doi.org/10.1145/1122445.1122456>

Developing Simple Web Text for People with Intellectual Disabilities and to Train Artificial Intelligence

Rochford, John ¹

*1. Eunice Kennedy Shriver Center, University of Massachusetts Medical School
55 Lake Avenue North, Worcester, Massachusetts 01655
John.Rochford@umassmed.edu*

RÉSUMÉ. Une mauvaise compréhension de la lecture pose des problèmes importants aux personnes souffrant de déficiences intellectuelles (PwID), personnes âgées, personnes dont la langue maternelle est différente de celle du texte ainsi qu'aux personnes peu alphabétisées. La simplification des textes vise à réduire la complexité du texte tout en conservant son sens. La recherche sur la simplification des textes pour PwID se poursuit depuis des décennies. Cependant, jusqu'à récemment, il n'y a pas eu d'effort significatif pour simplifier automatiquement le texte, spécifiquement pour PwID, en utilisant des technologies d'intelligence artificielle. Nous travaillons sur la simplification des textes sur le Web en utilisant des technologies d'intelligence artificielle (IA). La meilleure façon de déterminer l'efficacité des directives en langage clair que nous avons opérationnalisées est de faire évaluer le texte simplifié par des testeurs ciblés : des personnes présentant une déficience intellectuelle.

ABSTRACT. Poor reading comprehension presents significant challenges to people with intellectual disabilities (PwID); the aging; non-native language speakers; and people with low literacy. Text simplification aims to reduce text complexity while retaining its meaning. Text simplification research for PwID has been ongoing for decades. However, until recently, there has been no significant effort to automatically simplify text, specifically for PwID, using artificial intelligence technologies. In the short term, we are working to improve manual text simplification by creating We are working on Web text simplification using artificial intelligence (AI) technologies. The best way to determine the efficacy of the plain language guidelines we have operationalized is to have text simplified with them judged by expert testers: people with intellectual disabilities.

Mots-clés : simplification des textes, intelligence artificielle, trouble de la lecture.

KEYWORDS: text simplification, artificial intelligence, intellectual disabilities

1. Introduction

Text comprises most Web content. Poor reading comprehension presents significant challenges to people with intellectual disabilities (PwID). PwID have significant trouble with learning, reasoning, and problem-solving (*Intellectual and Developmental Disabilities (IDDs)*, 2021), and with everyday social and life skills. For PwID, difficulty understanding new information and comprehending abstract concepts are among common characteristics (*Inclusive Teaching: Intellectual Disability - ADCET*, 2014) that are significant barriers to understanding typical Web text.

Text simplification aims to reduce text complexity while retaining its meaning. Text simplification research for PwID has been ongoing for decades (Chen et al., 2016). However, until recently (Rochford, 2019), there has been no significant effort to automatically simplify text, specifically for PwID, using artificial intelligence (AI) technologies.

Our EasyText.AI Web text simplifier machine learning (ML) model (Rochford, 2019) uses a mix of AI technologies, including:

- neural machine translation that regards typical Web text sentences and simplified ones as distinct languages (Wang et al., 2016);
- eye-tracking as an objective measure of cognitive load / comprehension difficulty (Djamasbi, Shojaeizadeh, et al., 2016; Shojaeizadeh et al., 2017);
- development of operationalized plain-language guidelines (Djamasbi, Rochford, et al., 2016).

The best way to determine the efficacy of the plain-language guidelines we have operationalized is to have text simplified with them judged by expert testers: people with intellectual disabilities. Examples judged effective by PwID become part of the training data set for our EasyText.AI ML model.

2. Methodology

We augment our ML model with a generative adversarial network (GAN)⁹ based on operationalized plain language guidelines (OPLG) developed and proven effective by this pilot study and by other studies employing eye tracking (Djamasbi, Shojaeizadeh, et al., 2016; Shojaeizadeh et al., 2017). For our purposes, the GAN has two essential components: a generator of Web text examples using our OPLG and a discriminator that judges the examples using our OPLG.

For this pilot study, we created a subset of 6 plain language guidelines we operationalized by:

- 1) gathering 33 plain language guidelines from United States (*Plainlanguage.Gov | Home*, 2015; *WebAIM: Writing Clearly and Simply*,

2006) and international sources (*Home - Plain Language Association International (PLAIN)*, 2015);

- 2) narrowing them to 19 that could be quantified and defined by computer algorithms; and
- 3) conducting trials to determine the order they should be applied to Web text passages such that they significantly lowered grade reading levels, and such that all team members determined the same reading level results.

We obtained 8 exemplars of typical text from Websites used by people with intellectual disabilities (see Figure 2) (Rochford, 2010). Using the 6 operationalized plain-language guidelines, we rewrote 4 of the exemplars until they were at or lower than a United States 5th grade reading level. (see Figure 3.) Each text passage was short (about 25 to 50 words) and included only 2 concepts. For each passage, we wrote 2 literal and 2 inferential comprehension questions, 1 for each concept; and 3 multiple choice answers for each question. We assessed the reading levels of all passages and questions using standard readability tests (Friedman & Hoffman-Goetz, 2006). All questions and answers were at or below a 3rd grade level.

We recruited 25 adults with intellectual disabilities (ID) as participants (see Figure 1). We used a within-subject research design, so participants served as their own controls. This minimized mitigating factors such as differences in reading comprehension, IQ scores, and cognitive ability. We tested each participant's reading comprehension using the *Woodcock Johnson Reading Mastery Test* (Riverside Publishing, 2015). We gave participants regular breaks to minimize fatigue. We paid all participants for their time.

We presented each participant with 4 passages: 2 from the set of 4 typical-text passages (see Figure 2), and 2 from the set of 4 plain-language text passages (see Figure 3). We presented only 1 version of each passage to each participant. This prevented a comprehension bias that may have occurred if participants read typical-text passages and simplified versions of them. (Figures 2 and 3 are presented to demonstrate simplification. Both were not shown to participants.)

We presented participants with the text passages, resembling simple Webpages, on a personal computer monitor. Each Webpage displayed only the text of the passages, the questions, and the answers. We followed the same script for all sessions with participants. All communication between us and participants was verbal. We used verbal prompts exclusively. We, not participants, controlled the computer.

For each of the 4 passages, we presented 5 Webpages in the following sequence: text passage, then literal and inferential question and answer sets 1, 2, 3, and 4 (see Figures 3, 4, & 5). Each Webpage was shown by activating a (Next Page) button. We gave participants 5 minutes to read a text passage. At the start, we asked participants to read the text passage, more than once if needed, and indicate when they were ready to answer questions about it. At an affirmative answer, we presented them with question and answer set #1. At 4 minutes, 45 seconds, with a negative answer or without an answer, we advised them they had only a few seconds

before we showed them the first question; and at the five-minute mark, we presented the first question and answer set.



Figure 1. Recruitment

Sample Typical Passage Webpage

On Miniclip you can play games to compile games statistics such as high scores, and rankings on your own player page and can receive awards for playing games. You can also create your own Miniclip avatar with clothing and accessories to represent you in our fun Miniclip games. You can play single player games or play advanced massive multiplayer games with other players from around the world.

Figure 2. Sample Typical Text Passage

Sample Simplified Passage Webpage

On Miniclip, you can play games to collect game stats. These can be high scores and rankings on your own player page. You can get awards for playing games. You can also create your own Miniclip avatar with clothing and extras. It can be you in our fun Miniclip games. You can play single player games. Or you can play advanced, huge games with many players. Those players live around the world.

Figure 3. Sample Simplified Text Passage

Sample Typical Passage, Literal Question Webpage

Lux Level is a luxurious, in-theatre dining experience at select theatres. Movie-goers can indulge themselves with premium reserved-seating, in-seat dining throughout the show, as well as other special amenities. Each seat is equipped with a server call button so your server is always there when you need them. This truly is the finest movie-going experience available today.

Where can Lux Level be found?

1. At all theatres.
2. At a few theatres.
3. At no theatres yet.

Figure 4. Sample Typical Text Passage with Literal Question

Sample Simple Passage, Inferential Question Webpage

On Miniclip, you can play games to collect game stats. These can be high scores and rankings on your own player page. You can get awards for playing games. You can also create your own Miniclip avatar with clothing and extras. It can be you in our fun Miniclip games. You can play single player games. Or you can play advanced, huge games with many players. Those players live around the world.

How can you play with people who live far away?

1. Pick an advanced game.
2. Pick a single player game.
3. Pick an award.

Figure 5. Sample Typical Text Passage with Inferential Question

For each question and answer set, we asked participants to read the question and the answers and indicate the answer they thought correct. Once they told us their answer, we selected it on the Webpage (checked a checkbox) and asked if they were ready for the next question. We presented the next question and answer set at an affirmative answer, or after an answer change, we verified by asking if it was their final answer. Every participant answered every question.

3. Results

The primary goal of this within subject design pilot study (n=25) was to evaluate whether people with intellectual disabilities, when provided typical Web text simplified with plain language standards, demonstrated improved comprehension of the content. Additional related goals were to assess the possible relationship of age, baseline reading level testing scores, and process factors (including question type) on any comprehension improvement.

Socio-demographic data were limited to participant age. Sample age ranged from 17 to 63 years, with a mean chronological age of 38 years. Mean standard scores from this testing are in the 75-80 range (i.e. low) with age equivalents of 8-11 years and grade equivalents of 3rd to 5th grade.

The overall difference in percent correct on comprehension questions between the standard versus simplified passages was 18%, considered to be a large effect (Cohen's $d=1.1$). Note that the mean improvement for the literal questions also showed a large effect (Cohen's $d=1.2$) with a difference of 23%, while for the inferential questions the improvement was more modest at 13%, considered to be a medium effect (Cohen's $d=.6$)

Subsequent regression on these significant associations showed interesting predictive patterns. For each unit increase in the reading fluency standard score, the percent improvement decreased by 1.2%, suggesting as expected that those with the lowest reading literacy experience the most comprehension gain from simplification.

To examine this variation, percent correct on standard text comprehension tertiles were created to define 3 subgroups (very poor: a standard text comprehension score of < 60% correct, poor to fair: a standard text comprehension score of 60% - 79% correct, and good to very good: a pre-test level of 80% or above correct). The mean comprehension improvement scores for these three levels were 32.8% for the very poor group, 17% for the poor to fair group, and 0% for the good to very good group, with the group difference being statistically significant as shown by ANOVA ($p = .001$)

4.. Conclusion

A statistically significant comprehension difference between the standard (typical) passages and the simplified passages suggests the intervention is successful in improving comprehension through simplification. Benefit varied greatly as

expected based on comprehension of typical Website text, as well as based on reading fluency as measured by baseline reading testing. Effect sizes support the inferential results in suggesting large overall improvement, with improvement variation based on reading ability as well as the type of comprehension questions asked (literal versus inferential).

A post-hoc power analysis suggests that, with the overall large effect seen, the pilot sample size of 25 was more than sufficient for an overall paired difference comparison between standard versus simplified Website text comprehension, yielding power of 99%. A larger study, however, would enable additional subgroup analysis by other sociodemographic factors and on a more diverse population to evaluate the overall benefit of this simplification for other adults with cognitive impairments or with limited literacy skills.

Bibliographic

- Chen, P., Rochford, J., Kennedy, D. N., Djasmasbi, S., Fay, P., & Scott, W. (2016). Automatic Text Simplification for People with Intellectual Disabilities. In *Artificial Intelligence Science and Technology* (Vol. 1–0, pp. 725–731). WORLD SCIENTIFIC. https://doi.org/10.1142/9789813206823_0091
- Djasmasbi, S., Rochford, J., Daboll-Lavoie, A., Greff, T., Lally, J., & Mcavoy, K. (2016). Text Simplification and User Experience. *Proceedings, Part II, of the 10th International Conference on Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience - Volume 9744*, 285–295. https://doi.org/10.1007/978-3-319-39952-2_28
- Djasmasbi, S., Shojaeizadeh, M., Chen, P., & Rochford, J. (2016). Text Simplification and Generation Y: An Eye Tracking Study. *SIGHCI 2016 Proceedings*. <https://aisel.aisnet.org/sighci2016/12>
- Friedman, D. B., & Hoffman-Goetz, L. (2006). A systematic review of readability and comprehension instruments used for print and web-based cancer information. *Health Education & Behavior: The Official Publication of the Society for Public Health Education*, 33(3), 352–373. <https://doi.org/10.1177/1090198105277329>
- Home—Plain Language Association International (PLAIN). (2015). <https://plainlanguagenetwork.org/>
- Inclusive Teaching: Intellectual Disability—ADCET. (2014). <https://www.adcet.edu.au/inclusive-teaching/specific-disabilities/intellectual-disability>
- Intellectual and Developmental Disabilities (IDDs): Condition Information. (2021). <https://www.nichd.nih.gov/>. <https://www.nichd.nih.gov/health/topics/idds/conditioninfo/default>
- Plainlanguage.gov | Home. (2015). <https://www.plainlanguage.gov/>
- Riverside Publishing. (2015). *Riverside—Woodcock-Johnson III*. <http://www.riversidepublishing.com/products/wjIIIComplete/index.html>
- Rochford, J. (2010, February 12). *Web Accessibility Insights from 6 Women with Intellectual Disabilities*. <https://clearhelper.blog/2010/02/12/web-accessibility-insights-from-6-women-with-intellectual-disabilities/>

- Rochford, J. (2019). *Easy Text Artificial Intelligence*. Easy Text Artificial Intelligence. <https://easytext.ai/>
- Shojaeizadeh, M., Djasbi, S., Chen, P., & Rochford, J. (2017). Text Simplification and Pupillometry: An Exploratory Study. In D. D. Schmorow & C. M. Fidopiastis (Eds.), *Augmented Cognition. Enhancing Cognition and Behavior in Complex Human Environments* (pp. 65–77). Springer International Publishing. https://doi.org/10.1007/978-3-319-58625-0_5
- Wang, T., Chen, P., Rochford, J., & Qiang, J. (2016). Text simplification using Neural Machine Translation. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 4270–7271.
- WebAIM: Writing Clearly and Simply*. (2006). <https://webaim.org/techniques/writing/>

Atelier

SID et démocratie dans les organisations

4^{ème} Édition

Raphaëlle Bour, Maryse Salles

*IRIT, Université Toulouse 1 Capitole
rbour@irit.fr, Maryse.Salles@ut-capitole.fr*

Pour l'année 2021, l'atelier est axé sur des retours d'expériences et témoignages de professionnels et d'universitaires, qui peuvent nous éclairer sur l'impact des systèmes d'information et de décisions sur la démocratie des organisations.

Il y aura trois tables rondes (sur une demi-journée), qui nous offriront chacune un point de vue sur le sujet avec un entretien et un débat autour des invités suivants :

- un ou des professionnel, utilisateur de système d'information dans le cadre de son travail nous proposera son témoignage et son analyse à partir de son expérience « de terrain »,
- un universitaire, intéressé par l'analyse des systèmes d'information des organisations nous apportera son éclairage sur le sujet,
- un professionnel, acteur de la conception et du développement logiciel nous livrera sa vision du thème traité, et son retour d'expérience en tant que constructeur de système d'information.

Chaque table ronde débutera avec un entretien entre les invités et les membres de l'atelier, et sera suivie d'une discussion avec le public présent.

- **Table ronde 1**

Invités : deux utilisateurs de SI numériques (dans l'industrie et dans l'hôpital public) et deux universitaires (un socio-anthropologue, et un sociologue).

Partie 1 : Quel impact des SI numériques sur le travail ?

Partie 2 : Quel impact des SI numériques sur sens du travail, les valeurs des salariés et de l'organisation, la finalité de l'organisation ?

- **Table ronde 2**

Invités : Une universitaire (système d'information, agilité), un ingénieur d'étude (en charge de l'administration de systèmes numériques documentaires) et une coach agile d'une ESN.

Partie 1 : les projets informatiques permettent-ils une dispute démocratique ?

Partie 2 : les méthodes agiles sont-elles plus démocratiques que les méthodes de gestion de projet classiques ?

EcoSoft : Un éco-label pour la durabilité des logiciels

Rébecca Deneckère, Gregoria Rubio

*Centre de Recherche en Informatique
Université Paris 1 Panthéon-Sorbonne, Paris, France
Rebecca.Deneckere@univ-paris1.fr*

Cet article a été publié dans ISESL, l'atelier international sur l'ingénierie des systèmes d'information pour une vie plus intelligente (international workshop on Information Systems Engineering for Smarter Life), associé à la conférence internationale CAISE, en 2020 (Deneckère R, Rubio G. *EcoSoft: Proposition of an Eco-Label for Software Sustainability. Advanced Information Systems Engineering Workshops. 2020;382:121-132. Published 2020 Apr 29. doi:10.1007/978-3-030-49165-9_11*)

RESUME. On s'intéresse de plus en plus à la durabilité des entreprises et à la façon dont les entreprises devraient l'inclure pour répondre aux besoins des utilisateurs en ce qui concerne les impacts sociaux, économiques et environnementaux. Nous proposons un label écologique pour la durabilité des logiciels, basé sur des critères pertinents trouvés dans différents travaux.

Mots-clés : Durabilité, Ecolabel, Logiciel.

1. Introduction

La recherche sur la durabilité vise à offrir des méthodes, des techniques et des outils pour réduire l'impact des nouvelles technologies sur l'environnement, pour offrir un monde meilleur, une vie plus intelligente, aux générations futures. Les systèmes d'information doivent participer à l'effort collectif de développement durable, et les entreprises de logiciels et d'applications doivent mener des stratégies pour atteindre cet objectif. Au-delà d'une approche individuelle, la durabilité doit être considérée comme une qualité intégrale de tout logiciel. Dans cet article, nous nous sommes concentrées sur la durabilité environnementale. L'impact des nouvelles technologies sur l'environnement est assez ambigu car elles peuvent être à la fois bénéfiques et néfastes. Pour une analyse plus approfondie de leur impact, il est également nécessaire d'examiner les différentes phases de leur cycle de vie et les effets qui y sont associés. Il est important d'ajouter de la valeur aux logiciels développés de manière durable, et la labélisation est un moyen de leur donner plus de visibilité. Un

label de durabilité pour les logiciels et services numériques sert cet objectif d'apporter plus de visibilité et donc de crédibilité à l'écoconception dans le domaine des logiciels.

2. Travaux connexes

Les consommateurs sont de plus en plus nombreux à considérer la dimension écologique des produits qu'ils achètent et utilisent en fonction de l'intérêt croissant pour les questions environnementales (Eurobarometer, 2014). Des labels écologiques ont ensuite été introduits pour répondre à cette demande, avec plus de 450 labels de consommateurs dans le monde attribués par des gouvernements, des organisations ou des associations de consommateurs en 2016¹. Ces labels aident les consommateurs à identifier des produits plus responsables et à choisir selon des critères qui intègrent la qualité de l'environnement. Les labels, surtout si ils sont attribués par les autorités officielles, garantissent la confiance et la crédibilité des consommateurs (Darnal et Vasquez-Brust, 2016). Un éco-label est un label de qualité qui certifie qu'un produit ou un service a un impact réduit sur l'environnement. Il vise à « *promouvoir la conception, la production, la commercialisation et l'utilisation de produits ayant moins d'impact sur l'environnement tout au long de leur cycle de vie* » et « *mieux informer les consommateurs des effets des produits sur l'environnement, sans compromettre la sécurité du produit ou des travailleurs, ni affecter de manière significative les qualités qui rendent le produit adapté à son utilisation* »². Un label écologique peut certifier des produits et services de différentes catégories qui répondent à certains critères écologiques. Par exemple, nous trouvons des étiquettes pour l'alimentation (label Agriculture Biologique), le textile (label GOTS³), l'hébergement touristique (label Panda Gîte⁴), les appareils électriques (label Energy Star⁵), le bois (label FSC⁶) ou encore les cosmétiques (label Cos-mebio⁷).

Dans le domaine des nouvelles technologies, la labellisation est difficile à réaliser en raison de la dimension virtuelle des produits (logiciels, applications, sites web) qui fait croire aux utilisateurs qu'ils n'ont aucun impact sur l'environnement. Il est donc important d'ajouter de la valeur aux logiciels développés de manière durable, et la labellisation est un moyen de leur donner plus de visibilité. En s'engageant dans la mise en œuvre de certifications et de labels de technologies numériques, les gouvernements, en particulier aux niveaux européen et mondial, contribueraient à légitimer l'éco-conception de logiciels aux yeux des utilisateurs. Les entreprises qui s'efforcent de mettre en œuvre une stratégie durable dans le développement de leurs produits gagneraient également en crédibilité par rapport à leurs concurrents lorsque cette approche est officiellement reconnue (Darnall et Aragon-Correa, 2014). Un label

1 Ecolabel (2016) Index. All Ecolabels. <http://www.ecolabelindex.com/ecolabels/>

2 <http://www.ecolabels.fr/fr/espace-consommateurs/les-ecolabels>

3 <https://www.global-standard.org/fr/>

4 <https://ecotourisme.gites-de-france.com/notre-demarche-panda.html>

5 <https://www.energystar.gov/>

6 <https://fr.fsc.org/fr-fr>

7 <https://www.cosmebio.org/en/>

de durabilité pour les logiciels et services numériques tels qu'EcoSoft servirait donc cet objectif d'apporter plus de visibilité et donc de crédibilité à l'écoconception dans le domaine des logiciels. (Kern *et al.*, 2015) propose une discussion sur les labels des produits logiciels et des sites Web durables et énumère un ensemble de critères qui peuvent être appliqués. Les critères relatifs aux systèmes et aux applications Web de gestion de contenu Web écologique se trouvent dans (Dick *et al.*, 2012).

3. EcoSoft

Nous proposons ici l'établissement d'un label de durabilité pour les logiciels qui permettrait aux utilisateurs de choisir des programmes et des applications en fonction de leur impact sur l'environnement. Cette étiquette permettrait aux consommateurs d'identifier facilement le logiciel le plus durable et ainsi de limiter leur utilisation de carbone en tant qu'utilisateurs. C'est aussi une approche qui permet de sensibiliser l'entreprise aux impacts environnementaux des logiciels et services numériques et de mettre en œuvre un engagement en faveur du développement durable. L'entreprise réduit son empreinte logicielle tout en promouvant son image à travers cette approche d'éco-conception. Tout comme le label Energy Star sur les appareils électroniques économes en énergie, le label EcoSoft servirait à informer sur l'engagement volontaire pris par les concepteurs sur l'efficacité énergétique des logiciels sur un ordinateur, une tablette ou un smartphone selon des critères généraux. Il serait valable pour une période de 2 ans, durée maximale compte tenu des nombreux développements et mises à jour qui se produisent dans la vie d'un logiciel ou d'une application. Le logiciel ou l'application Web se verrait attribuer un logo distinctif pour rendre visible sa certification. Il existe un projet similaire pour l'éco-conception de sites Web en particulier appelé Green Code Label. Toutefois, aucun label européen ou international ne certifie la durabilité d'un service d'application. En effet, l'éco-label officiel de l'Union européenne ne certifie pas les produits et services numériques, ce qui ferait du label EcoSoft une référence dans le domaine de la certification de logiciels durables. Un logiciel étiqueté EcoSoft doit répondre à des critères de durabilité. Il doit avoir été développé conformément aux techniques d'écoconception de logiciels en intégrant les outils et les méthodes de gestion durable du cycle de vie complet du produit. EcoSoft est un label multi-critères, c'est-à-dire que son attribution est basée sur le respect d'un ensemble de critères prédéfinis, et ne peut donc pas être accordé sur la base d'un seul critère. Les logiciels, applications et services numériques de toutes catégories et de tous types peuvent être certifiés. Il n'y a pas non plus de restriction quant à la source du programme.

Selon (Naumann *et al.*, 2011), un logiciel peut être analysé selon les trois phases de son cycle de vie (développement, utilisation et fin de vie). Notre label EcoSoft utilise des critères classés en fonction de ces trois phases (cf. tableau 1). Cependant il n'aborde que les critères spécifiques concernant les logiciels et ne prend pas en compte les critères relatifs à la consommation d'énergie des ordinateurs et autres matériels ainsi que l'énergie nécessaire à l'infrastructure de l'entreprise. Nous avons choisi des critères dans les deux facettes de la durabilité définies par (Calero et Piattini, 2015) : la durabilité PAR le logiciel (ICT4Green) et la durabilité DANS le logiciel (GreenIT). L'attribution du label devra signifier que tous les critères de la liste sont respectés.

Tableau 1. EcoSoft Criteria

Development phase	
Sustainable documents	Download size
Sustainable specification	Hardware requirements
Code optimization	Green analysis efficiency
Usage phase	
Accessibility	Material resource consumption
Usability	Backup size
Energy efficiency at running time	
End of life phase	
Data conversion to the future	Long-term data storage
Software or application	Packaging and manuals

4. Conclusion

EcoSoft est un éco-label prenant en compte la participation des parties prenantes (chef de projet, architectes logiciels, développeurs) dans le processus d'intégration de la durabilité dans le projet. Les trois étapes du cycle de vie du logiciel (développement, utilisation et fin de vie), sont analysées pour déterminer les impacts environnementaux générés. Nous avons concentré l'analyse sur la consommation d'énergie des composants logiciels, aspect important pour la qualité globale du logiciel, en particulier pour l'expérience utilisateur sur l'appareil mobile, mais aussi car la consommation d'énergie numérique a une empreinte environnementale élevée.

Bibliographie

- Calero C., Piattini M. (2015) Green in software engineering. Springer International Publishing
- Darnall, N., Aragon-Correa, JA. (2014) Can ecolabels influence firms sustainability strategy and stakeholder behaviors ?. *Organization & Environment*, 27(4), pp 319–327
- Darnall, N., Ji, H., Vazquez-Brust, D. T (2016) hird-Party Certification, Sponsorship, and Consumers' Ecolabel Use, *J Business Ethics*, Vol. 150, Issue 4, pp 953–969
- Dick, M., Kern, E., Johann, T., Naumann, S., and Gulden, C. (2012). Green Web Engineering - Measurements and Findings. 26th International Conference on Informatics on Informatics - Informatics for Environmental Protection, Sustainable Development and Risk Management (EnviroInfo); Federal Environment Agency, Dessau
- Eurobarometer (2014) Attitudes of European citizens towards the environment, Special Eurobarometer. http://ec.europa.eu/public_opinion/archives/ebs/ebs_416_en.pdf
- Kern, E., Dick, M., Naumann, S., Filler A. (2015) Labelling Sustainable Software Products and Websites: Ideas, Approaches, and Challenges. 29th International Conference on Informatics for Environmental Protection (EnviroInfo), Third International Conference on ICT for Sustainability (ICT4S)
- Naumann, S., Dick, M., Kern, E., Johann, T. (2011) The GREENSOFT Model: A reference model for green and sustainable software and its engineering, *Sustainable Computing: Informatics and Systems*, 1(4), 294–304

Fouille de Processus de Configuration dans les Lignes de Produits

Camille SALINESI¹, Housseem CHEMINGUI^{1,2}, Inès GAM^{1,2}, Raul MAZO³, Henda BEN GHEZALA²

- 1. Centre de Recherche en Informatique, Université Paris 1 Panthéon Sorbonne, France.*
- 2. Laboratoire RIADI, Ecole Nationale des Sciences de l'Informatique, Université la Manouba, Tunisie.*
- 3. Laboratoire STICC, École Nationale Supérieure de Techniques Avancées Bretagne, France.*

Le présent article fait référence à une publication discutée dans la conférence CENTERIS - International Conference on ENTERprise Information Systems, organisée à Sousse, Tunisie du 16 au 18 Octobre 2019. La référence bibliographique complète de l'article en style APA est reprise ci-dessous :

Chemingui, H., Gam, I., Mazo, R., Salinesi, C., & Ben Ghezala, H.. (2019). Product Line Configuration Meets Process Mining. *Procedia Computer Science*, 164, 199-210. <https://doi.org/10.1016/j.procs.2019.12.173>

1. Contexte et Problématique

Les lignes de produits (Pohl et al., 2005) présentent une stratégie de production qui consiste à développer les éléments d'une famille de produits similaires par réutilisation. Cette démarche offre aux entreprises un avantage concurrentiel en améliorant la productivité et la qualité, tout en favorisant une meilleure adéquation aux besoins via la personnalisation.

Cependant, les avantages potentiels des lignes de produits peuvent être perdus lorsque les modèles de variabilité qui permettent de guider la réutilisation sont de grande taille et complexes. En effet, de tels modèles impliquant des centaines de variantes conduisent à une explosion combinatoire du nombre de combinaisons possibles. De très nombreuses actions de personnalisation peuvent être effectuées, chaque action de personnalisation pouvant produire des résultats très différents. Dans ce cadre, les activités de configuration deviennent sujettes aux erreurs, et il devient très difficile voire impossible de tenir compte de toutes exigences et des contraintes qu'il faudrait satisfaire simultanément. Ce verrou fondamental ne peut être résolu qu'en offrant aux parties prenantes un guidage des processus de configuration complexes.

2. Etude de l'existant

L'étude exploratoire des problèmes de la configuration des lignes de produits, détaillée dans l'article référencé, montre une grande similitude entre les problèmes signalés dans la littérature, entre autres, l'incompréhension des choix, la pléthore des variantes, la lourdeur d'exécution des décisions, etc. En effet, certains articles utilisent des termes différents pour exposer le même problème. Par conséquent, trois classes de problèmes ont été identifiés, notamment : (i) la complexité de la décision, (ii) le manque de performance et (iii) le besoin de guidage pour naviguer dans la collection de décisions possibles.

Face à ces problèmes, différentes approches ont été rapportées, par la communauté des lignes de produits, pour guider les parties prenantes, entre autres, les configurations dirigées par des problèmes de satisfaction de contraintes (Salinesi et al., 2011), les configurations dirigées par les heuristiques (Mazo et al., 2014) et les configurations guidées par les recommandations du filtrage collaboratif (Triki et al., 2014).

Les solutions existantes pallient les problèmes de configuration en une partie. Néanmoins, ces solutions s'avèrent caduques face à des parties prenantes de plus en plus exigeantes et différentes. En effet, ces solutions produisent du guidage rigide, standard et parfois dépendant à l'arborescence du modèle de la ligne de produits. De plus, la considération des buts et des différences de la partie prenante demeure omise. Ensuite, les approches recensées recommandent des décisions ou des instances de décisions mais pas les deux avec un ordre de processus de configuration. Les problèmes identifiés et le besoin de recommander des valeurs et un ordre de processus en même temps correspondent amplement aux apports de la fouille de processus (Van der Aalst, 2012).

3. Proposition et motivations

L'article référencé explore et discute la contribution d'une recommandation basée sur la fouille de processus pour surmonter les problèmes de configuration. L'idée est non seulement d'aider les parties prenantes dans les arbres de décision, mais aussi de les aider à choisir les "meilleures" alternatives de configuration. Pour ce faire, une approche nouvelle a été proposée consistant à tracer les actions des parties prenantes, reconstituer les modèles de processus correspondants et y tirer des connaissances pour les prochaines expériences.

Le guidage du processus de configuration est fondé sur le "Fuzzy Miner" (Günther et al., 2007), un algorithme de fouille de processus recommandé pour la découverte et l'analyse de processus présentant des comportements non structurés. Dans le processus de configuration de lignes de produits, les décisions des parties prenantes sont conflictuelles, interdépendantes et susceptibles, souvent, d'influencer négativement le processus entier. En effet le "Fuzzy Miner" permet d'avoir une représentation graphique et dynamique des décisions du processus de configuration en mettant en valeur l'ordre suivi par l'acteur de la configuration et les

comportements habituels pour atteindre une configuration donnée. Pour ce faire, le “Fuzzy Miner” applique diverses techniques, telles que la suppression de tronçons sans importance, le clustering de nœuds fortement corrélés dans un seul nœud et la suppression des clusters de nœuds isolés. De surcroît, l’interprétation des modèles de processus de configuration découverts via le “Fuzzy Miner” revêt deux dimensions, notamment : une concentration sur l’ordre de processus des décisions et une idée plus claire sur les instances de décisions.

Pour parvenir à produire du guidage fondé sur la fouille, les modèles de processus de configurations antérieures sont interrogés et interprétés suivant des buts de guidage notamment (i) la performance, (ii) la personnalisation et (iii) la flexibilité de la configuration. Afin d’atteindre un but de guidage, sélectionné généralement par la partie prenante, un processus de recommandation est conduit pour réutiliser les modèles de processus découverts et orchestrer leur apports tout au long de processus de guidage. Pour ce faire, les décisions qui auront un impact en faveur de l’atteinte du but de guidage en question seront recommandées. La mesure de l’impact des décisions de configuration est calculée suivant des directives de recommandation (Schonenberg et al., 2008) à savoir l’abstraction des réponses de guidage, leur pondération et le calcul de leurs aboutissements.

Sur le plan technique, une architecture multicouche comprenant une couche de présentation, une couche de fouille et une couche de recommandation, a été conçue. Les résultats des expérimentations préliminaires montrent une objectivité et une transparence des modèles de processus découverts outre la capacité à accélérer les processus de configuration et identifier les goulots d’étranglement des processus. Par ailleurs, la fouille de processus de configuration peut aller plus loin que suggérer une valeur pour une instance de décision de configuration, soit la sélection ou le refus d’une variante du modèle de variabilité, mais aussi de suggérer un ordre des décisions de configuration. La recommandation basée sur la fouille de processus plutôt que la fouille de données statiques, peut révéler des informations sur le comportement d’exécution et l’enchaînement des étapes de configuration. Ainsi, proposer aux parties prenantes des formes de guidage personnalisées et dirigées par des buts, incluant des instances des décisions dans un ordre dynamique s’avère avantageux.

3. Limites et Perspectives

Plusieurs questions scientifiques se posent concernant les conséquences et les inconvénients de cette forme de guidage pour les praticiens. En effet, lorsque le modèle de variabilité évolue, l’écart entre ce qui est capitalisé dans les traces et ce que contient le nouveau modèle est inévitable. Par conséquent, le guidage devient incohérent avec les nouvelles évolutions. En outre, le phénomène de démarrage à froid constitue aussi une pierre d’achoppement pour le guidage fondé sur la fouille de processus. En effet, cette dernière n’a de sens que si un nombre important de traces est déjà capitalisé.

Par analogie aux lignes de produits, les autres systèmes configurables tels que les (ERP) et les (COTS) sont aussi contraints par le grand nombre d'exigences volatiles et de contraintes à satisfaire simultanément. La configuration devient rapidement sujette à l'erreur et confronte les parties prenantes à des problèmes de rigueur et de passage à l'échelle lors de la configuration de logiciels : les méthodes employées pour spécifier les systèmes à configurer ne permettent pas de maîtriser de manière formelle et systématique un nombre important de décisions complexes. Cependant, des approches visant à surmonter ces verrous existent et ont été publiées, mais dans certains domaines; peuvent-elles être adaptées pour toute sorte de logiciels configurables? Les défis scientifiques de la configuration logicielle peuvent-ils ainsi être transposés? (Chemingui et al., 2020)

References

- Chemingui, H., Gam, I., Mazo, R., Ghezala, H. B., & Salinesi, C. (2020). A Unified Vision of Configurable Software. In *INFORSID* (pp. 93-98).
- Günther, C. W., & Van Der Aalst, W. M. (2007). Fuzzy mining–adaptive process simplification based on multi-perspective metrics. In *International conference on business process management* (pp. 328-343). Springer, Berlin, Heidelberg.
- Mazo, R., Dumitrescu, C., Salinesi, C., & Diaz, D. (2014). Recommendation heuristics for improving product line configuration processes. In *Recommendation Systems in Software Engineering* (pp. 511-537). Springer, Berlin, Heidelberg.
- Pohl, K., Böckle, G., & van Der Linden, F. J. (2005). *Software product line engineering: foundations, principles and techniques*. Springer Science & Business Media.
- Salinesi, C., Mazo, R., Djebbi, O., Diaz, D., & Lora-Michiels, A. (2011). Constraints: The core of product line engineering. In *2011 fifth international Conference on Research Challenges in Information Science* (pp. 1-10). IEEE.
- Schonenberg, H., Weber, B., Van Dongen, B., & Van der Aalst, W. (2008, September). Supporting flexible processes through recommendations based on history. In *International Conference on Business Process Management* (pp. 51-66). Springer, Berlin, Heidelberg.
- Triki, R., Mazo, R., & Salinesi, C. (2014). Combining configuration and recommendation to enable an interactive guidance of product line configuration.
- Van Der Aalst, W. (2012). Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS)*, 3(2), 1-17.

Le Machine Learning pour un Context Mining Facility

Nourhène BEN RABAH^{1,2}, Manuele KIRSCH PINHEIRO¹,
Bénédicte LE GRAND¹, Ali JAFFAL², Carine SOUVEYET¹

1. Centre de Recherche en Informatique, Université Paris 1 Panthéon Sorbonne
90 rue de Tolbiac, 75013 Paris, France
{Manuele.Kirsch-Pinheiro|Benedicte.Le-Grand|Carine.Souveyet}@univ-paris1.fr
2. ESIEE-IT, 8 rue Pierre de Coubertin I 95 300 Pontoise, France
{nbenrabah|ajaffal}@esiee-it.fr

RÉSUMÉ. *Le Machine Learning (ML) est de plus en plus utilisé dans les applications sensibles au contexte pour la fouille des données de contexte. Dans l'article (Ben Rabah et al., 2020), nous posons la question de la généralisation de la fouille à l'échelle de tout un Système d'Information (SI), en tant que « facilité » proposée à l'ensemble de ses composants. Ce changement d'échelle pose de nombreux défis à l'application des techniques de ML, notamment pour le prétraitement de données et l'optimisation des hyperparamètres. L'étude de littérature présentée dans cet article invite à une prise de conscience indispensable pour l'avènement d'un service de « context facility » dans les SI.*

MOTS-CLÉS : *Machine Learning, Fouille de contexte, Systèmes d'Information*

1. Introduction

Les informations de contexte sont de plus en plus aisées à capturer, offrant de multiples opportunités au sein des Systèmes d'Information (SI) (adaptation au contexte, recommandation de services ou de contenu, prédiction des besoins ou des actions des utilisateurs, etc.), surtout lorsqu'elles sont combinées à des techniques de *Machine Learning (ML)*. De nombreuses applications sensibles au contexte utilisent ainsi ces techniques (Ameyed, 2015). On peut alors envisager la généralisation de ces comportements « *intelligents* », en proposant la fouille de contexte comme un service propre à un SI.

Nous avons proposé dans (Ben Rabah *et al.*, 2020) une vision d'une « *context mining facility* », *i.e.*, un service de fouille de contexte généralisé à tout un SI, ainsi utilisable par toutes ses applications. Un tel service représenterait un changement majeur d'échelle dans la fouille de contexte, habituellement réalisée en tenant compte d'une application précise, avec des besoins précis en termes de données et des techniques d'apprentissage, comme par exemple dans (Yuan & Herbert, 2014) (Mokhtari *et al.*, 2017) (Rueda et al. 2019). L'hétérogénéité des données de contexte

apparaît alors comme un obstacle majeur pour leur analyse généralisée et totalement automatisée. Dans (Ben Rabah *et al.*, 2020), nous avons analysé l'utilisation du ML pour la fouille de contexte et les conditions nécessaires pour que cette analyse puisse tenir compte des particularités propres aux informations de contexte, ainsi que les défis pour sa généralisation à l'échelle d'un SI. Nous les résumons ici pour la communauté SI française.

2. Le *Machine Learning* pour une *context facility*

La généralisation de la fouille de contexte à l'échelle d'un SI représente la possibilité d'offrir, à l'échelle de tout un système, des mécanismes de raisonnement disponibles en continu pour toutes les applications et évoluant avec le système, avec un minimum d'intervention humaine. Ce support doit être pensé pour tout un système, avec son écosystème d'utilisateurs (décideurs, employés, clients...) et d'applications supportant les différents processus et corps de métiers qui cohabitent dans une organisation. Un même service de fouille de contexte devrait être capable de supporter les besoins de multiples applications (existantes ou futures), en intégrant à tout moment de nouvelles informations de contexte hétérogènes et en constante évolution. Dans ces conditions, on ne peut pas avoir de connaissance préalable de ces données.

Avec des mots clés tels que *context prediction / awareness, ML, data mining* et Internet of Things (*IoT*), nous avons identifié 200 articles et étudié 30 en détail. À partir de ces articles, nous avons pu identifier des pratiques propres aux techniques de ML que nous avons confronté aux caractéristiques propres aux informations de contexte également soulignées dans la littérature.

Nous avons observé que, quel que soit le paradigme adopté (apprentissage supervisé, non supervisé, demi-supervisé ou par renforcement), le ML consiste à sélectionner un algorithme et à l'entraîner sur certaines données. L'efficacité d'une telle méthode dépend de divers facteurs, tels que la qualité des données d'entraînement, l'algorithme choisi et de ses hyperparamètres. Des données de mauvaise qualité peuvent compromettre le succès des algorithmes ML les plus puissants. Garcia et al. (2016) et Ramirez-Gallego et al. (2017) reconnaissent d'ailleurs l'importance de la qualité des données. Or les données de contexte sont, par définition, hétérogènes et incertaines (Bettini et al., 2010) (Chalmers et al. 2004). Souvent issues de différents capteurs, ces données brutes peuvent être erronées et incomplètes, ce qui peut entraîner un impact négatif sur le temps d'entraînement, la complexité et la qualité des modèles générés par le ML. De même, certaines techniques de ML requièrent une distribution équilibrée des données dans les classes disponibles, ce qui n'est pas toujours possible avec les données de contexte, en raison du caractère dynamique et incertain de ces données. Une phase de prétraitement de données comprenant plusieurs étapes (le nettoyage, la transformation, l'extraction et la sélection et l'extraction des caractéristiques, l'augmentation, etc.) et une éventuelle intervention d'un expert humain sont alors nécessaires pour régler ce problème de qualité des données récoltées et pouvoir les exploiter.

Par conséquent, l'introduction de nouvelles données dans le système implique l'arrêt du service pour réexécuter le prétraitement de ces données et l'entraînement de(s) modèle(s) sélectionné(s). Si on considère un service de « *context facility* » généralisé à l'ensemble d'un SI, un tel arrêt pourrait impacter d'autres applications potentiellement critiques pour l'organisation. Il paraît alors clair qu'en état les techniques de ML actuelles ne semblent pas encore assez adaptées pour un usage à l'échelle de tout un SI.

3. Conclusions

La réussite des techniques de ML repose encore beaucoup sur l'expertise humaine pour le choix de l'algorithme le plus approprié, l'optimisation des hyperparamètres, la préparation des données, etc. Or une application généralisée de ces techniques dans le cadre de tout un SI, tel que nous envisageons dans (Ben Rabah *et al.*, 2020), demanderait davantage d'automatisation afin de permettre son usage par des multiples applications et afin de minimiser les temps d'arrêt pour ce service. Il existe néanmoins de plus en plus de *framework* et d'outils, comme Automated Machine Learning (AutoML) (Hutter *et al.*, 2019) qui simplifient l'usage du ML, voire automatisent partiellement l'exécution de leur pipeline. On peut donc espérer que les difficultés identifiées dans (Ben Rabah *et al.*, 2020) pourront s'amenuiser avec le temps, pour que cette vision d'un « *context mining facility* » puisse voir le jour. Il y en va du même pas pour l'explicabilité de ces techniques, qui reste un facteur important pour l'acceptabilité de celles-ci auprès des usagers.

Bibliographie

- Ameyed, D., Miraoui, M., Tadj, C. (2015). A survey of prediction approach in pervasive computing. *International Journal of Scientific & Engineering Research*, vol. 6, n° 5, p. 306-316.
- Ben Rabah, N., Kirsch Pinheiro, M., Le Grand, B., Jaffal, A., Souveyet, C. (2020). Machine Learning for a Context Mining Facility. *16th Workshop on Context and Activity Modeling and Recognition*, 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), p.678-684.
- Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, vol. 6, n° 2, Apr 2010, p. 161-180.
- Chalmers, D., Dulay, N., Sloman, M. (2004). Towards Reasoning About Context in the Presence of Uncertainty. *1st Int. workshop on advanced context modelling, reasoning and management*. Nottingham, UK, September 2004.
- García, S., Luengo, J., Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. *Knowledge-Based Systems*, vol. 98, 15 April 2016, p. 1-29.
- Hutter, F., Kotthoff, L., Vanschoren, J. (2019). *Automated Machine Learning: Methods, Systems, Challenges*, Springer, p. 219. Doi: 10.1007/978-3-030-05318-5.

- Mokhtari, G., Zhang, Q., Fazlollahi, A. (2017). Non-wearable UWB sensor to detect falls in smart home environment. CoMoRea 2017, IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), p. 274-278.
- Ramírez-Gallego, S., Krawczyk, B., García, S., Woźniak M. , Herrera, F. (2017). A survey on data preprocessing for data stream mining: Current status and future directions. *Neurocomputing*, vol. 239, 24 May 2017, p. 39-57.
- Rueda, F. M., Lüdtke, S. Schröder, M., Yordanova, K., Kirste, T., Fink, G. A. (2019). Combining Symbolic Reasoning and Deep Learning for Human Activity Recognition. CoMoRea 2019, IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), p. 22-27.
- Yuan B., Herbert, J. (2014). Context-aware Hybrid Reasoning Framework for Pervasive Healthcare. *Personal Ubiquitous Computing*, vol. 18, no. 4, p. 865-881.

Author Index

A	
Aiello, Antoine	13
Antoine-Santoni, Thierry	13
Araujo, David	13
B	
Ben Ghezala, Henda	103
Bour, Raphaëlle	96
C	
Cardinale, Yudith	31
Cardon, Remi	61
Cayèré, Cécile	40
Chemingui, Houssein	103
D	
Dafflon, Baudouin	49
Delhom, Marielle	13
Deneckère, Rébecca	99
E	
Ermakova, Liana	56
G	
Gam, Ines	103
Ghzaïel, Imene	49
González, Mauricio Jacobo	31
Grabar, Natalia	61
Gualtieri, Jean-Sébastien	13
I	
Ilarri, Sergio	31
J	
Jaffal, Ali	107
K	
Kahn, Giacomo	49
Kirsch Pinheiro, Manuele	1
Kirsch-Pinheiro, Manuele	3, 107
L	
Laborie, Sébastien	31
Larracochea, Jorge Andrés	31
Laval, Jannik	49
Le Grand, Benedicte	107
M	
Mameli, Alexandre	13
Mazo, Raul	103
Mccombie-Boudry, Helen	80
Mondo, Melanie	40

Mothe, Josiane	56
N	
Niang, Boubou Thiam	49
Nourhène Ben Rabah, Nourhène	107
O	
Ouzrout, Yacine	49
P	
Poggi, Bastien	13
Ponsard, Christophe	21
R	
Radia, Hannachi	64
Riahi, Kenza	49
Richard, Jérémy	40
Rochford, John	88
Roose, Philippe	1, 31
Rubio, Gregoria	99
S	
Salinesi, Camille	103
Salles, Maryse	96
Sanjuan, Eric	56
Simeoni, Jean-Marc	13
Souveyet, Carine	1, 3, 107
Steffenel, Luiz Angelo	1
Sílvia, Araújo	64
T	
Touzani, Mounir	21
U	
Unwalla, Mike	75
V	
Van Hieu, Ho	13
Vittori, Evelyne	13