



HAL
open science

Extraction de connaissances à partir de l'usage des outils professionnels de communication et de collaboration

Antoine Flepp, Fabrice Bourge, Julie Dugdale, Tiphaine Marie

► To cite this version:

Antoine Flepp, Fabrice Bourge, Julie Dugdale, Tiphaine Marie. Extraction de connaissances à partir de l'usage des outils professionnels de communication et de collaboration. Rencontres des Jeunes Chercheurs en Intelligence Artificielle (RJCIA 2017), Jul 2017, Caen, France. hal-01561981

HAL Id: hal-01561981

<https://hal.science/hal-01561981v1>

Submitted on 2 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Extraction de connaissances à partir de l'usage des outils professionnels de communication et de collaboration

Antoine Flepp^{1,2}

Fabrice Bourge¹

Julie Dugdale^{2,3}

Tiphaine Marie¹

¹Orange Labs – France

²Univ. Grenoble Alpes, CNRS, Grenoble INP*, LIG, F-38000 Grenoble France

³Université de Adger – Norvège

Orange Labs P&S
42 rue des Coutures – BP 6243
14066 Caen Cedex 4
antoine.flepp@orange.com

Résumé

La multiplication des outils de communication et de collaboration (média sociaux, e-mail, conférence à distance, etc.) a entraîné ces dernières années une explosion du volume de données non-structurées. Or, ces données (contenus et usages des outils) recèlent une connaissance explicite et tacite de valeur, souvent inexploitée. Cette connaissance apporte un contexte essentiel pour comprendre les documents de référence existants – plus structurés – qui ont été produits de manière collaborative, permettant à tout employé de donner un sens à leur contenu et favoriser ainsi la collaboration. Après avoir décrit différentes classifications de connaissances, les outils de communication et de collaboration sont catégorisés sur la base de travaux précédents dans le domaine du travail collaboratif, puis au travers de théories sur les médias. Une approche bottom-up est proposée pour extraire la connaissance en utilisant des méthodes quantitatives et des méthodes qualitatives (i.e. des entretiens, questionnaires, pop-ups, fichiers de traces, etc.). Cette connaissance sera utilisée dans le but de développer un modèle qui servira de base pour des collaborations plus efficaces et plus efficaces.

Mots Clef

Outils de communication et de collaboration, connaissance explicite, connaissance tacite, connaissance contextuelle, sense-making, CSCW.

Abstract

The multiplication of communication and collaboration tools (social media, e-mail, conferencing, etc.) in recent years has led to an explosion of the quantity of unstructured data. Buried within this data – both content and the way we use these tools – lies valuable explicit and tacit knowledge, often untapped. This knowledge provides essential context for understanding the existing, more

structured reference documents that were produced in a collaborative manner, allowing people to make sense of their contents and further aiding collaboration. After discussing various classifications of knowledge, communication and collaboration tools are categorised based on previous work in the domain of collaborative work as well as media theories. A bottom-up approach is proposed to extract the knowledge using both quantitative and qualitative methods (i.e. interviews, questionnaires, pop-ups, log files, etc.). This knowledge will be used in order to develop a model that will serve as a basis for more efficient and effective collaborations.

Keywords

Communication and collaboration tools, explicit knowledge, tacit knowledge, contextual knowledge, sense-making, CSCW.

1 Introduction

Chaque jour la grande majorité des nouvelles données générées et stockées au sein des entreprises sont non-structurées, c'est-à-dire qu'elles ne sont pas organisées selon des modèles prédéfinis. Comme on constate depuis de nombreuses années une croissance de la multiplicité des outils de communication et de collaboration (C&C) – e.g. les outils de messagerie, de conférence à distance, les média sociaux – il est raisonnable de penser qu'une partie significative de ces données non structurées réside dans ces outils : courriels, documents, blogs, contenus audio ou vidéo, etc. Or, ces données constituent potentiellement un gisement prodigieux d'information et de connaissance encore très peu exploité aujourd'hui. Aussi nous avons l'ambition de proposer un cadre de recherche exploratoire visant à extraire, formaliser et modéliser la connaissance non seulement à partir des contenus manipulés ou échangés via ces outils de C&C, mais aussi à partir des usages de ceux-ci. Au-delà de l'intérêt pratique au sein des entreprises, l'apport scientifique est d'extraire et de rendre exploitable des connaissances tacites et latentes.

² Institute of Engineering Univ. Grenoble Alpes

Nos travaux ayant débuté récemment, cet article se concentre principalement sur un état de l'art. Ainsi, la partie 2 présente une synthèse de plusieurs travaux sur les connaissances tacites et explicites, afin de définir un cadre d'analyse sur le type de connaissance que les données extraites pourront apporter. Ce cadre doit permettre *in fine* de modéliser et de distinguer les connaissances obtenues par les données de contenu, de celles obtenues par les données d'usage. Ensuite nous présentons diverses études sur les outils de C&C, afin d'identifier les données pertinentes qui peuvent être extraites à partir de ces outils. Il y est ainsi décrit le caractère formel ou informel des échanges que ces outils permettent, des caractéristiques importantes de classifications des outils, puis est abordé le potentiel de ces outils à pouvoir compléter ou remplacer les échanges en face-à-face. Dans la partie 3 nous proposons notre hypothèse de travail. Cette dernière stipule qu'à partir du recueil et de l'analyse des données (de contenus et d'usages) issues des outils de C&C – utilisés lors d'un travail collaboratif – nous devrions être en mesure d'en extraire de la connaissance contextuelle (i.e. relative au contexte du travail collaboratif), pour ensuite rendre cette connaissance utile et accessible pour toute personne extérieure à ce contexte. La partie 4 précisera l'approche suivie, ainsi que les éléments de méthodologie associés. Enfin, dans la conclusion, nous évoquerons brièvement le contenu de cet article ainsi que les étapes futures de ces travaux de recherches.

2 Contexte

2.1 Classifications de la connaissance

De nombreuses études se sont succédées suite aux travaux de Michael Polanyi sur les dimensions tacite et explicite de la connaissance [24]. Deux types d'approches semblent être adoptés : une approche descriptive qui tend à éclaircir les ambiguïtés entre les concepts de connaissance tacite et explicite [15,20], et une approche opérationnelle cherchant plutôt à décrire les processus par lesquels on passe de connaissances tacites à des connaissances explicites et inversement [6,7,13,23]. Ainsi, Polanyi [24] a d'abord décrit les connaissances tacites et les connaissances explicites en deux catégories distinctes. Les connaissances explicites seraient celles codifiées sous une forme structurée (e.g. graphique, tableur, base de données, rapport d'activité). Alors que les connaissances tacites seraient les connaissances non encore codifiées ni structurées que les individus possèdent dans leur cerveau et qui s'acquerraient en majeure partie par l'expérience, l'observation ou encore l'échange d'idées [20]. De fait, la connaissance explicite semble facilement – et à moindre coût – accessible à un grand nombre de personnes, mais elle ne représente probablement qu'une partie de l'ensemble de la connaissance [24]. Ensuite, ce concept de connaissance tacite et de connaissance explicite a été repris – en termes de processus – au travers du modèle SECI de Nonaka et Takeuchi [23], et plus récemment dans le modèle conceptuel de PKM 2.0 de Cheong et

Tsuis [6]. Plus précisément, le modèle SECI – qui tire son nom de : Socialisation, Externalisation, Combinaison, Internalisation – décrit l'interaction entre connaissances tacites et explicites, tant individuelles que collectives, comme un processus en spirale. Le modèle de Cheong, quant à lui, se focalise sur les connaissances individuelles uniquement, d'où le nom de PKM pour *Personal Knowledge Management* (gestion des connaissances personnelles). Globalement, l'auteur propose dans son modèle quatre composants principaux : *Personal Information Management* (gestion des informations personnelles), *Personal Knowledge Internalisation* (internalisation des connaissances personnelles), *Personal Wisdom Creation* (création d'une sagesse personnelle) et *Inter-Personal Knowledge Transferring* (transfert de connaissances interpersonnelles). De fait, ces modèles décrivent les connaissances tacites et les connaissances explicites en termes de processus individuel ou collectif, plutôt qu'en termes de catégories distinctes. C'est pourquoi des auteurs comme Jasimuddin [15] suggèrent finalement qu'il faut concevoir ce concept de connaissance comme un continuum dont la connaissance tacite et la connaissance explicite en seraient les deux extrêmes. Ceci se confirme et s'étaye par des travaux [7,13] qui se poursuivent sur l'articulation des connaissances individuelles et collectives avec les connaissances tacites et explicites. En outre, Grundstein [13] propose que les connaissances individuelles tacites (e.g. talents, habiletés, expérience) amèneraient à des connaissances collectives tacites ; comme, par exemple, des routines défensives (qui feraient obstacles au changement) et des routines offensives (qui favoriseraient l'innovation et le changement). Plus récemment, Collins [7] distingue la connaissance tacite relationnelle (les choses que l'on peut décrire en principe si on s'en donne la peine), la connaissance tacite somatique (les choses que notre corps est capable de faire mais que nous ne pouvons pas décrire) et la connaissance tacite collective (la connaissance que nous tirons du fait de faire partie de la société, comme les règles de langage). De fait, aucun consensus n'apparaît quant à ces différents types de connaissances (tacite individuelle / tacite collective / explicite individuelle / explicite collective). Cependant, il est important de soulever un point que ces deux derniers auteurs considèrent : la connaissance tacite peut être explicitable ou non (*explicable knowledge*). Et c'est en particulier Collins [7 : p.81] qui traite de ce propos, en distinguant différents processus d'explicitation de la connaissance : par élaboration, par transformation, par mécanisation ou par explication (voire par *induction*). Ainsi, même si comme Lejeune [20] l'affirme : « *Dès qu'il est formalisé, le savoir n'est plus tacite, il s'est transformé* » ; il s'avère – au vu des recherches présentées – qu'il y a des connaissances tacites plus difficiles à formaliser que d'autres (*explicable knowledge*), et qu'à en croire Collins, il existe aussi différentes manières de les formuler.

2.2 Les outils de communication et de collaboration

Au fil du temps, l'entreprise est passée d'une organisation hiérarchique à une organisation fonctionnelle caractérisée par une formalisation de ses règles de fonctionnement. Ce modèle vise une interdépendance la plus faible possible entre les entités spécialisées qui la composent afin, notamment, de limiter les flux considérés comme un mal nécessaire ; chaque entité s'inscrit dans la réalisation de processus définis [3]. Néanmoins, du fait de la complexification de l'environnement de travail et de la multiplication des rôles assumés par les individus, Bertin et Tran suggèrent que l'on vit actuellement une phase de transition vers un modèle multipolaire dans lequel l'entreprise est composée d'entités de plus en plus multifonctionnelles. Celles-ci tendent à s'auto-organiser pour réaliser des activités pour des groupes – plus ou moins temporaires – qui ne dépendent pas nécessairement des mêmes entités hiérarchiques. Une des conséquences est l'augmentation des flux transverses d'information et de connaissance, ainsi que la nécessité de collaborer davantage en dehors des processus définis, ce qui induit moins de formalisme. Le foisonnement des flux de communication et de coordination est renforcé par la multiplication des outils de C&C qui semble suivre un « effet millefeuille » [17], à savoir un empilement successif des outils les uns sur les autres, sans réorganisation des processus de gestion associés.

De nombreuses classifications des outils de C&C existent de par le monde. En 1988, dans le cadre du CSCW (Computer-Supported Cooperative Work) Johansen [16] propose de catégoriser les outils selon leur contexte d'utilisation via une matrice spatio-temporelle permettant de distinguer si un travail collectif est réalisé en un seul ou plusieurs lieux géographiques, et en même temps (synchrone) ou à des moments différents (asynchrone). Diverses extensions ont été proposées depuis, par exemple, pour prendre en compte la notion de processus : Andriessen [2] retient les processus liés aux échanges entre personnes, aux tâches et aux activités de groupe ; alors que Grudin et Poltrock [12] préfèrent distinguer les processus de communication, de partage d'information et de coordination. Plus récemment, Alimam et al. [1] proposent une typologie de divers outils de collaboration sociale utilisés en entreprise tels que les principaux outils de télécommunication, les outils de collaboration (CSCW, groupware...) et les outils sociaux (Entreprise 2.0). Après l'évaluation de ces outils au regard de treize critères, une typologie est produite en retenant trois des critères indépendants les plus significatifs : diffusion de l'information (modèle expéditeur-destinataire(s) ou modèle *publish-subscribe* – littéralement, publier-s'abonner) ; conception de l'interface utilisateur (classique – i.e. conforme à un modèle communément utilisé ce qui en facilite l'usage et l'acceptabilité – ou spécifique – dont l'usage et l'acceptabilité sont difficiles à

évaluer) ; fonction principale de l'outil (i.e. communication, partage ou recherche d'information, édition ou organisation). La première catégorie de cette typologie regroupe les outils qui permettent une interaction selon le modèle expéditeur-destinataire(s), dont la conception est classique, et qui partagent la même fonction principale de communication et de partage d'information. Ces outils, que les auteurs qualifient de canaux classiques de communication privée, sont : le téléphone, la messagerie vocale, le *chat*, le SMS et l'e-mail. La seconde catégorie de cette typologie regroupe les outils qui suivent le modèle *publish-subscribe*. On distingue alors deux sous-catégories. La première rassemble les outils de conception classique ; leur fonction principale concerne l'organisation (marque-page social) et la gestion de l'information (moteur de recherche, flux d'activité et RSS, streaming / wiki / blog / micro-blog, tableau blanc, forum / communauté / groupe de discussion). La seconde sous-catégorie regroupe les outils de conception spécifique ; leurs fonctions principales sont celles de communication et de partage d'information (conférence à distance), de gestion d'information (partage de fichier, recherche sociale, planification collaborative, banque d'idées, profil social) ou d'organisation (agenda et gestion de calendrier, suivi et localisation, gestion de contenu, application composite (*mash-up* en anglais), application métier).

Outre ces typologies, il existe aussi la théorie de la richesse des médias – proposée par Daft et Lengel [9]. Cette dernière décrit les outils de C&C selon leur potentiel à transmettre une information riche entre des interlocuteurs (e.g. le téléphone ne transmet pas les expressions du visage, au contraire de la visioconférence). Cette théorie a ensuite été approfondie à plusieurs reprises. On trouve par exemple la théorie de la naturalité des médias introduite par Kock [19], qui stipule que le potentiel d'un média à reproduire l'information transmise dépendrait de son degré de similarité avec une communication face-à-face. L'auteur en conclut qu'il y aurait cinq éléments qui décrivent la naturalité d'un média : la co-implantation (i.e. être au même endroit), la synchronicité (i.e. au même moment) et la capacité à transmettre des expressions langagières, faciales et corporelles. Ainsi, le manque de naturalité d'un média entraînerait chez les interlocuteurs plus d'ambiguïté dans leur communication, un effort cognitif supplémentaire, ainsi qu'une survenue plus rapide de l'ennui dans l'interaction. L'auteur a pu confirmer – avec des études empiriques [18] – ces effets négatifs d'un manque de naturalité d'un média. Néanmoins, cette même étude a aussi montré que la qualité de la tâche demandée n'avait pas été affectée, laissant suggérer ce que l'auteur appelle une adaptation compensatoire (*compensatory adaptation*). Cependant, on trouve aussi d'autres alternatives à la théorie de la richesse des médias, telles que la théorie de la synchronicité des médias de Dennis [10]. Cette théorie – contrairement à la *naturalité des médias* – suggère que

la communication face-à-face n'est pas toujours le médium le plus « riche », mais que le meilleur média ou ensemble de médias dépend de cinq dimensions selon une situation donnée, qui sont : l'immédiateté de la rétroaction, le parallélisme, la variété des symboles, la capacité à peaufiner un message avant de l'envoyer (*rehearsability*) et à rejouer plusieurs fois un message après réception (*reprocessability*). Concrètement, par exemple, le téléphone a un haut degré d'immédiateté de la rétroaction mais un faible degré de *reprocessability* contrairement au courriel. Outre ces théories sur différentes typologies et potentiels des outils de C&C, on trouve aussi des théories centrées *utilisateur*. Par exemple, selon Haythornthwaite [14], plus des pairs auront des liens forts (e.g. liens sociaux ou professionnels) et plus ils utiliseront de médias de communication. Au contraire, des pairs qui auraient des liens faibles, utiliseraient alors peu de médias pour communiquer entre eux. Néanmoins, cette théorie peut être nuancée selon les propos de Conein [8] qui suggèrent qu'« à partir d'un certain seuil un groupement n'est plus viable et se fragmente. Les contraintes cognitives concernent nos capacités limitées de traitement des informations qui portent sur des relations sociales ». Enfin, on peut aussi citer la théorie de l'expansion des canaux de Carlson et Zmud [5], qui aborde notre perception des canaux de communications selon l'expérience utilisateur. Les auteurs proposent quatre types d'expériences utilisateurs pertinentes : expérience du canal de transmission, expérience du sujet du message, expérience du contexte organisationnel et expérience avec les coparticipants. Ainsi, comme l'ont en partie suggéré Kock et les auteurs de cette dernière étude, les utilisateurs s'adaptent aux médias de manière compensatoire. En d'autres termes, avec une plus grande expérience dans l'un des quatre domaines décrits par Carlson et Zmud (canal, sujet, contexte, partenaires), les utilisateurs de médias auraient une meilleure capacité de communiquer dans divers contextes situationnels et percevront ainsi le canal utilisé comme étant plus riche.

3 Propositions

Nous tenons à préciser que nous serons amenés dans le futur à distinguer plusieurs types de connaissances tacites et explicites. Pour autant, dans cette première phase de nos recherches, nous considérons uniquement l'hypothèse selon laquelle les contenus échangés avec les outils de C&C sont plutôt sources de connaissance explicite (donc plus facilement extractible) et les usages, plutôt sources de connaissance tacite (plus difficilement extractible). Les outils de C&C semblent incontournables pour échanger et collaborer au sein d'une entreprise. Ainsi, le travail individuel pour la production de données structurées (e.g. un *document de référence* : communiqué de presse, livrable) tend à faire place au travail collaboratif. En outre, comme l'illustre la figure 1 avec la zone de *Production* (en pointillé), le *travail collaboratif* qui se faisait principalement en face-à-face entre des

contributeurs peut se faire – à présent – de plus en plus au travers des outils de C&C (bulle orange). Par ailleurs, des études comme celles de Yuan [25] ont montré que souvent les *documents de référence* (correspondant à de la connaissance explicite), ne se suffisent pas à eux-mêmes pour qu'ils soient compréhensibles par le plus grand nombre. Concrètement, lors de l'exploitation de *documents de référence* par des *non-contributeurs* (N-C), ces derniers se retrouvent confrontés à un manque d'informations pour appréhender ces *documents de référence* et doivent contacter directement un *contributeur* du document, voire en discuter avec d'autres *non-contributeurs*, pour pouvoir achever leur processus de compréhension. Au sein d'Orange, par exemple, cette situation se présente à chaque fois qu'une personne intègre un projet déjà en cours, et qu'elle doit s'approprier le contexte au travers de la lecture des documents déjà produits par certains groupes de travail. De même, tout nouvel arrivant dans l'entreprise (apprenti, stagiaire...) est confronté à cette situation puisqu'il doit monter en compétence le plus rapidement possible via la lecture de multiples documents. On peut noter que ce processus de compréhension – qui sera à approfondir dans les recherches futures – est appelé *sense-making* pour des auteurs comme Maurel [21] ou *brainfiltering* par d'autres [4]. Notre hypothèse de travail est que les informations transmises via les outils de C&C – que ce soit au niveau des *contributeurs* entre eux, entre *contributeurs* et *non-contributeurs* ou *a posteriori* au niveau des *non-contributeurs* entre eux – peuvent être automatiquement détectables et exploitables pour en extraire de la connaissance qui sera utile au sein de toute l'entreprise. De plus, différentes études [11,19,22,25] démontrent l'importance du contexte dans la recherche d'informations et de connaissances. Et bien que des auteurs aient mis l'accent sur certaines propriétés contextuelles de la connaissance comme les différences culturelles [25], la complémentarité des outils (e.g. usage de la messagerie

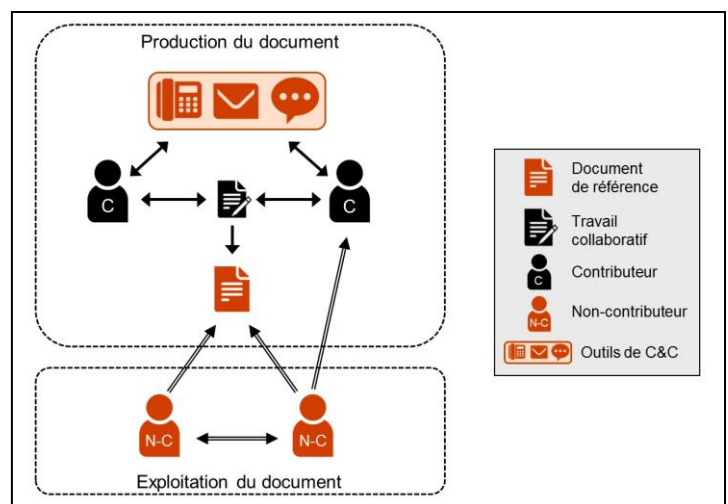


Figure 1: Modélisation d'extraction de connaissances à partir des outils de C&C

instantanée si la personne à contacter est en ligne / usage du téléphone si c'est urgent / usage du courriel si ce n'est pas urgent [22]), le contexte social [19], etc. ; il sera sans doute possible de considérer, dans nos futurs travaux, la *connaissance contextuelle* comme facteur de l'*awareness of expertise distribution* de Yuan [25], voire plus généralement de la *situation awareness* [11].

4 Méthodologie

Une approche *top-down* consisterait d'abord à modéliser toute la connaissance d'une entreprise puis à essayer de valider ce modèle sur le terrain. Cependant, il peut y avoir incompatibilité entre le modèle et les données du terrain, d'où une validation difficile. C'est pourquoi nous préférons suivre une approche *bottom-up*, plus pragmatique se décomposant en trois phases : la collecte des données, l'analyse de ces données et la transformation des résultats en connaissance (production d'un modèle de représentation de connaissance). Dans cet article, on se focalise sur la première phase.

4.1 Outils ciblés

Il est important d'étudier l'usage de plusieurs outils en parallèle car, d'une part, ceux-ci présentent différentes formes d'interaction (plus ou moins intrusive, formelle/informelle, synchrone/asynchrone, etc.). D'autre part, les personnes utilisent ces outils souvent de manière combinée pour répondre à leur besoin de partage de la connaissance. Ainsi, les outils que nous ciblons a priori pour la collecte de données sont ceux qui sont les plus utilisés au sein d'Orange. Trois de ces outils correspondent au modèle expéditeur-destinataire(s) : l'e-mail, la messagerie instantanée, la conférence à distance. Et un quatrième outil, mis en œuvre plus récemment, correspond au modèle *publish-subscribe* : le réseau social d'entreprise.

4.2 Données ciblées

Comme indiqué dans la partie 3, nous souhaitons analyser les contenus échangés et partagés lors d'un travail collaboratif, ainsi que les données d'usages de ces outils de C&C utilisés pour ce travail. Par contenus, nous entendons les contenus proprement dit (i.e. ce qui est manipulé par les utilisateurs) et les données qui les accompagnent (e.g. nom de l'émetteur, date d'envoi). Les données d'usage, quant à elles, correspondent aux informations relatives aux actions des utilisateurs sur les outils (e.g. envoyer un message, ouvrir un outil, chercher un contact). C'est la complémentarité de ces deux types de données qui devrait nous permettre d'extraire la connaissance contextuelle.

4.3 Mise en œuvre

Pour collecter des données d'usage, nous mènerons dès que possible des entretiens et soumettrons des questionnaires auprès de quelques personnes qui collaborent fréquemment (soit parce qu'elles font partie de la même équipe, soit parce qu'elles contribuent au même projet) de manière à favoriser la dimension de

partage de connaissances. Lors de cette étape, nous collecterons également auprès de ces mêmes personnes des contenus qu'ils auront générés ou manipulés. Parallèlement, nous récupérerons des fichiers de traces au niveau des serveurs internes de l'entreprise quand ce sera possible. Aussi nous devrions disposer rapidement d'un ensemble de données porteuses de valeur de par leur diversité et complémentarité. Le croisement de ces contenus avec les données d'usages devrait nous permettre d'en déduire des informations quant à l'activité des utilisateurs en fonction des contenus échangés, des personnes avec qui ils interagissent et du moment de la journée. L'objectif est de comprendre le pourquoi et le comment des actions de collaboration. En outre, ceci devrait nous permettre d'affiner la liste des outils qui avaient été ciblés a priori.

Dans un deuxième temps, il nous semble important d'automatiser la collecte des données afin de disposer de volumes significatifs. Ainsi, nous développerons un logiciel – installé sur le poste de travail – pour récupérer certains contenus échangés et leurs données associées, et pour générer des fichiers de traces en local. L'intérêt de ces derniers est de pouvoir tracer finement la cohérence des différentes actions des utilisateurs sur l'ensemble de leurs outils, ce qui est difficile à obtenir à partir des seuls fichiers de traces au niveau des serveurs. Par la suite, nous ferons évoluer ce logiciel afin de recueillir auprès des utilisateurs des données qualitatives sur leurs activités collaboratives. Le recueil de ces avis doit être fait au fil de l'eau, afin que l'utilisateur soit le plus spontané et objectif possible. Ceci de manière à éviter ainsi d'éventuels biais dus au délai et au changement d'état d'esprit entre l'action et l'avis que l'on y porte. Une étape de validation de cet outil sera mise en œuvre afin, notamment, de s'assurer qu'il ne soit pas trop intrusif.

4.4. Évaluation des travaux

Il nous semble difficile de pouvoir évaluer nos travaux à l'issue de cette première phase de collecte des données dans la mesure où celles-ci seront porteuses de valeur latente uniquement. Néanmoins, les différentes techniques mises en œuvre ont pour but de maximiser les chances de collecter des données porteuses de valeur pour les phases suivantes.

C'est une fois qu'on aura transformé en connaissances les résultats de l'analyse des données collectées qu'il nous sera possible d'évaluer nos travaux. Moyennant d'identifier les méthodes théoriques ou expérimentales adaptées, deux pistes s'offrent à nous : d'une part évaluer la pertinence et la complémentarité des informations venant « enrichir » un document de référence et, d'autre part, mesurer quantitativement et qualitativement le gain d'efficacité et d'efficience d'un utilisateur lorsqu'il bénéficie d'informations complémentaires pour l'aider à comprendre des documents de référence.

5 Discussions

Dans la suite de nos travaux, il sera nécessaire de comparer et d'appliquer différentes méthodes d'extraction de connaissances. Ces dernières devront répondre à différents points importants d'évaluation, tels que la valeur de la connaissance extraite, non seulement du point de vue de leur pertinence, mais aussi en termes de fidélité par rapport aux échanges lors de la collaboration via les outils de C&C. Compte tenu de l'approche bottom-up proposée, les techniques sémantiques de traitement automatiquement de la langue naturelle – et plus précisément terminologiques – nous semblent a priori plus adaptées pour répondre à notre problématique, bien qu'elles seront sans doute à combiner avec des méthodes statistiques.

6 Conclusion

Les outils de C&C font de plus en plus partie intégrante de notre activité professionnelle quotidienne. C'est pourquoi leurs contenus et leurs usages constituent – pour l'entreprise – une source d'information de grande valeur, encore peu exploitée jusqu'à présent. De fait, nous proposons d'extraire de la connaissance à partir du contenu des outils, ainsi qu'à partir des données de leurs usages. Plus précisément, notre objectif est de générer de la connaissance relative au contexte d'un travail collaboratif (*connaissance contextuelle*) – réalisé via des outils de C&C – afin d'en faire bénéficier tout collaborateur extérieur. L'état de l'art présenté permet d'obtenir un premier cadre exploratoire de recherche, que ce soit sur les connaissances, les outils de C&C ou encore la relation qu'il peut y avoir entre la connaissance et ces outils, à savoir que : les données de contenus des outils de C&C seront sources de connaissances explicites et les données d'usages de ces outils, sources de connaissances tacites. Ainsi, nos travaux se porteront dans un premier temps sur la collecte des données qui ont de la valeur a priori à travers les contenus et les usages des outils de C&C. Puis, les étapes suivantes – que nous présenterons à l'occasion d'articles futurs – porteront à la fois sur la mise en œuvre d'outils pour analyser les données collectées, ainsi que sur la formalisation et la modélisation de la connaissance ainsi générée.

Bibliographie

- [1] M. Alimam, E. Bertin et N. Crespi, "Enterprise Social Systems: The what, the why, and the how," in *17th IEEE Conf. Bus. Informatics (CBI)*, 2015, pp. 9-17.
- [2] J.H.E. Andriessen, *Working with groupware: understanding and evaluating collaboration technology*, Springer Science & Business Media, 2003.
- [3] E. Bertin et S. Tran, "La ville, le SI l'entreprise: du fonctionnel au multipolaire," *Management & Avenir*, 2014, n° 2, pp. 54-72.
- [4] E. Benitez et D. Pauleen, "Brainfiltering: The Missing Link Between PKM and PIM ?," dans *Amer. Conf. on Inform. Syst. (AMCIS) 2009 Proc.*, 13.
- [5] J. R. Carlson et R. W. Zmud, "Channel expansion theory and the experiential nature of media richness perceptions," in *Academy of management journal*, 1999, vol. 42, n° 2, pp. 153-170.
- [6] R. Cheong et E. Tsui, "From Skills and Competencies to Outcome-based Collaborative Work : Tracking a Decade's Development of Personal Knowledge Management (PKM) Models," *Knowledge and Process Management*, 2011, vol. 18, n° 3, pp. 175-193.
- [7] H. Collins, *Tacit and explicit knowledge*, University of Chicago Press, 2010.
- [8] B. Conein, "Cognition distribuée, groupe social et technologie cognitive," *Réseaux*, 2004, n° 2, pp. 53-79.
- [9] R.L. Daft et R.H. Lengel, "Organizational information requirements, media richness and structural design," *Management science*, 1986, vol. 32, n° 5, pp. 554-571.
- [10] A. R. Dennis et J. S. Valacich, "Rethinking media richness: Towards a theory of media synchronicity," in *Proc. 32nd Annu. Hawaii Int. Conf. on Syst. Sci.*, IEEE, 1999.
- [11] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 1995, vol. 37, n° 1, pp. 32-64.
- [12] J. Grudin et S. Poltrock, "Computer Supported Cooperative Work," in *The Encyclopedia of Human-Computer Interaction*, 2nd Ed., Denmark, 2013.
- [13] M. Grundstein. De la capitalisation des connaissances au management des connaissances dans l'entreprise, les fondamentaux du knowledge management. *INT-Entreprises*, 2013, vol. 3, p. 256-272.
- [14] C. Haythornthwaite, "Social networks and Internet connectivity effects," *Information, Community & Society*, 2005, vol. 8, n° 2, pp. 125-147.
- [15] S. M. Jasimuddin, J. H. Klein, et C. Connell, "The paradox of using tacit and explicit knowledge: strategies to face dilemmas," *Management decision*, 2005, vol. 43, n° 1, pp. 102-112.
- [16] R. Johansen, *Groupware: Computer support for business teams*, The Free Press, 1988.
- [17] M. Kalika, N. B. Charki et H. Isaac, "La théorie du millefeuille et l'usage des TIC dans l'entreprise," *Revue française de gestion*, 2007, n° 3, pp. 117-129.
- [18] N. Kock, "Compensatory adaptation to media obstacles: An experimental study of process redesign dyads," *Information resources management journal*, 2005, vol. 18, n° 2, p. 41.
- [19] N. Kock, "Media richness or media naturalness? The evolution of our biological communication apparatus and its influence on our behavior toward e-communication tools," *IEEE Trans. on Prof. Commun.*, 2005, vol. 48, no 2, p. 117-130.
- [20] M. Lejeune, "Transmettre l'informel en milieu de travail : proximité, posture et modalités organisationnelles," dans *Formations et professionnalisations : à l'épreuve de la complexité*, Paris, L'Harmattan, 2012, chapitre V, pp. 107-115.
- [21] D. Maurel, "Sense-making : un modèle de construction de la réalité et d'appréhension de l'information par les individus et les groupes," *Études de communication*, 2010, n° 2, pp. 31-46.
- [22] T. Niinimäki, A. Piri, C. Lassenius et al, "Reflecting the choice and usage of communication tools in GSD projects with media synchronicity theory," in : *5th IEEE Int. Conf. Global Software Eng. (ICGSE)*, 2010, pp. 3-12.
- [23] I. Nonaka et H. Takeuchi, *The knowledge creating company: how Japanese companies create the dynamics of innovation*, New York : Oxford University Press, 1995.
- [24] M. Polanyi. *The tacit dimension*. First published Doubleday & Co, 1966.
- [25] C. Yuan, X. Zhao, Q. Liao, et al, "The use of different information and communication technologies to support knowledge sharing in organizations: From e-mail to micro-blogging," *Journal of the American Society for Information Science and Technology*, 2013, vol. 64, n° 8, p. 1659-1670.