Du traçage de l'activité informatique à sa redocumentation en texte

Une approche narrative et un outil auteur

Leila Yahiaoui[1,2], **Yannick Prié**[2], **Zizette Boufaida**[1]

1. *Université de constantine*
   Université Mentouri, laboratoire LIRE, Constantine 25000, Algérie
   {leila.yahiaoui, zboufaida}@gmail.com

2. *Université de Lyon, CNRS*
   Université Lyon 1, laboratoire LIRIS, UMP5205, F-69622, France
   {leila.yahiaoui, yannick.priei}@liris.cnrs.fr

RÉSUMÉ. *Dans ce travail, nous présentons une approche générique pour le processus de redocumentation de l'activité médiée informatiquement. Cette approche combine le contenu de traces de l'activité et les connaissances des utilisateurs dans des formes narratives textuelles, facilement échangeables, décrivant l'activité de façon personnalisée pour les utilisateurs. Elle utilise des modèles formels pour les traces et les documents produits, ainsi que les principes de la théorie de la structure rhétorique (RST). Elle soutient la redocumentation à travers deux phases : une phase automatique pour générer un document texte initial et une phase interactive pour personnaliser ce document. Nous présentons notre outil de redocumentation ActRedoc, une implémentation de notre approche, ainsi que les résultats de sa première évaluation.*

ABSTRACT. *In this paper, we present a generic approach for the redocumentation process of computer-mediated activity. This approach combines traces' content and users' knowledge in textual narrative forms, easily exchangeable, describing the activity in personalised way for users. It uses formal models for traces and the produced documents, as well as principles of Rhetorical Structure Theory (RST). It supports the redocumentation through two phases: an automatic phase to generate an initial text document and an interactive phase to personalize this document. We describe our redocumentation tool ActRedoc, an implementation of our approach, as well as the results of its first evaluation.*

MOTS-CLÉS : *trace d'activité informatique, modèle de trace, modèle de document, processus de redocumentation, RST, génération du langage naturel, outil auteur.*

KEYWORDS: *computer-mediated activity trace, trace model, document model, redocumentation process, RST, natural language generation, authoring tool.*

DOI:10.3166/TSI.31.1-15 © 2012 Lavoisier, Paris

1. Introduction

La documentation, en tant que processus, permet d'enregistrer et de collecter des informations pour fournir une preuve de quelque chose selon un objectif particulier, ce qui implique à la fois la création et la gestion de documents. Or, la technologie numérique a fait perdre aux documents leur stabilité comme objets matériels et des bouleversements majeurs se sont imposés sur leur production et leurs échanges. Dans ce contexte, Roger Pédauque (Pédauque, 2006) a souligné l'apparition d'un nouvel ordre documentaire, caractérisé par un processus dit de « redocumentarisation ». Son but est de créer ou de s'approprier des documents existants selon une vision personnelle. Aujourd'hui, l'émergence des technologies du *lifelogging* fait la preuve d'une telle tendance en offrant la possibilité de capturer les expériences de la vie quotidienne d'un individu à travers des artéfacts numériques ; entrainant de nouvelles opportunités de raconter et de partager des expériences via la création de narrations (Byrne et Jones, 2009).

Nous nous intéressons à la *documentation* de l'activité informatique *à partir de ses traces.* Nous entendons par « *activité informatique* » toute activité initiée par un humain et supportée par un système informatique permettant l'acquisition, le traitement, le stockage et/ou l'échange de données dans le but de réaliser l'activité en question. Cela inclue les activités de bureautique, la navigation sur le web, le *e-learning*, l'utilisation des applications en ligne ou des systèmes informatiques plus ou moins complexes, etc. La documentation de cette activité vise à fournir une description persistante, facilement échangeable et personnalisée (selon les choix d'un individu [1]) de l'activité en question. Le document produit est intéressant puisqu'il permet de revoir l'activité de l'utilisateur tracé et de réfléchir à son propos ; comme il peut servir à communiquer et à échanger des expériences personnelles. Les activités informatiques peuvent être enregistrées automatiquement à travers l'instrumentation des systèmes en vue d'enregistrer et de collecter des événements qualifiés de pertinents. Nous appelons traces les produits de ces enregistrements (fichiers logs, captures d'écran, enregistrements vidéo, etc.) qui fournissent une documentation de base pour l'activité tracée. Néanmoins, ces traces manquent le plus souvent de lisibilité, d'intelligibilité et de portée sociale pour des utilisateurs ordinaires (au contraire des concepteurs des systèmes, par exemple). De plus, ces traces sont souvent de nature multimodale et réparties sur plusieurs fichiers. Cela incite à concevoir des processus de *redocumentation* de l'activité informatique pour pouvoir transformer et corréler les traces de celle-ci au sein d'unités cohérentes et compréhensibles décrivant l'activité.

Différents projets et outils – tels que Oracle Trace Analyser [2], Cyclope Employee Monitoring Software [3] et (Gray et *al.*, 2004) – ont adopté une approche de

---

1. L'individu qui effectue ce processus peut être l'utilisateur qui a effectué l'activité informatique tracée (et que nous appelons *utilisateur tracé)* ou tout autre personne.
2. Oracle Trace Analyzer: http//www.oracleutilities.com/Packages/TraceAnalyzer_IOUG.pdf

redocumentation automatique. Leur principe est de générer des rapports organisés (hypertextes) à partir des traces d'activités. Cependant, cette approche est restrictive pour des utilisateurs qui préfèrent s'approprier le contenu et la forme des traces selon des choix personnels. D'autres projets utilisent la modélisation des traces (Champin et *al.*, 2003; Settouti et *al.*, 2009) pour proposer des transformations basées-modèles et des visualisation graphiques assurant un haut niveau d'abstraction pour représenter au mieux l'activité. La modélisation sémantique des traces améliore certes leur expressivité et facilite leur exploitation automatique, mais il reste que des informations importantes sur le contexte de l'activité tracée ne peuvent être collectées automatiquement ou modélisées explicitement (par exemple, l'opinion de l'utilisateur tracé).

Par conséquent, nous pensons que l'utilisateur doit pouvoir intervenir durant le processus de redocumentation de l'activité tracée pour décider concrètement du contenu à préserver des traces, du contenu à ajouter, et de la forme d'appropriation du document produit. En revanche, étant donnée la complexité potentielle des traces, leur volume et leur nature multimodale, il est inconcevable de demander à l'utilisateur de construire tout seul un document décrivant l'activité directement à partir de ses traces. Il est nécessaire de lui fournir un outil assistant qui permette de minimiser son effort cognitif durant le processus de redocumentation. De plus, nous proposons à l'utilisateur d'adopter une approche naturelle dans la description de l'activité, basée sur la construction de narration (récit). En effet, la narration est identifiée comme un élément central dans la manière dont les humains apprennent à donner du sens à leur monde (Nelson, 1989) et à leurs expériences.

Dans ce travail, nous présentons une approche semi-automatique pour la redocumentation en texte de l'activité informatique. Notre cadre général d'implémentation du processus, ses traces d'entrée et ses documents de sortie sont fondés sur des modèles (ontologies). Deux phases sont définies pour le processus : une phase automatique qui permet de générer un document texte fragmenté à partir des traces comme description initiale de l'activité, et une phase interactive qui permet à l'utilisateur de s'approprier la forme et le contenu de la description initiale de l'activité en adoptant une approche narrative basée sur les principes de la théorie de la structure rhétorique (RST) (Mann et Thompson, 1988). Nous décrivons ActRedoc, notre outil générique de redocumentation, ainsi que son évaluation initiale. La suite du papier est organisée comme suit : dans la section 2 nous présentons notre vue générale du processus de redocumentation de l'activité informatique, la section 3 présente un état de l'art sur la (re)documentation de l'activité, la section 4 est dédiée à la description de notre cadre général pour le processus de redocumentation en texte de l'activité informatique à partir de ses traces modélisées, la section 5 présente l'exemple d'une activité informatique à redocumenter, la section 6 décrit l'outil de redocumentation ActRedoc et la section 7 présente une première évaluation de son utilisation.

[3.] Cyclope Employee Monitoring Software : http://www.cyclope-series.com/employee-surveillance/employee-monitoring-software.html.

2. Une vue générale du processus de redocumentation de l'activité informatique

Le processus de redocumentation de l'activité informatique peut impliquer l'utilisateur effectuant l'activité (utilisateur tracé) ou toute autre personne (amis ou collègues de l'utilisateur tracé, analystes, psychologues, etc.). Ce processus permet d'exploiter les traces d'activité informatique, générées automatiquement, afin de les reprendre au sein d'un nouveau document numérique décrivant l'activité tracée selon l'intention et les choix de celui qui redocumente. Par conséquent, redocumenter implique *réorganiser, filtrer et réécrire le contenu des traces, y ajouter de nouvelles informations* (pour refléter un point de vue particulier ou pour exprimer le contexte de production des traces), ainsi que *choisir une forme d'appropriation pour le document produit* (Yahiaoui et *al.*, 2008).

La figure 1 présente notre vue globale du processus de redocumentation. L'entrée du processus est un ensemble de traces d'activité résultantes de l'enregistrement de l'interaction de l'utilisateur tracé avec le système informatique médiatisant son activité. Sa sortie est un document numérique décrivant l'activité tracée de façon personnalisée et échangeable. L'individu concerné par le processus peut garder le document produit à titre personnel (archivé), le destiner à ses proches (envoyé par mail ou via les réseaux sociaux à la famille, à des amis ou à des collègues), ou le rendre accessible au grand public (publié sur le web).

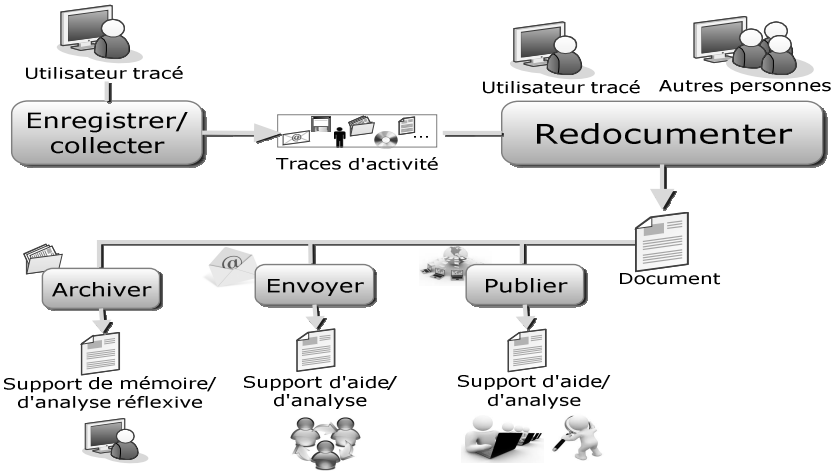


Figure 1. Vue globale du processus de redocumentation de l'activité informatique.

La fonction du document produit dépend de l'objectif de la redocumentation. Lorsque l'utilisateur tracé effectue la redocumentation de son activité, il peut conserver le document produit comme un support de mémoire ou d'analyse réflexive à propos de son activité. Il peut également échanger ce document décrivant son expérience avec d'autres personnes afin de leur servir de support d'aide à propos

de l'activité (pour des utilisateurs ayant une activité similaire, par exemple), ou encore de support d'analyse à propos de l'activité ou du comportement de l'utilisateur tracé (pour des analystes ou des psychologues, par exemple). Le document produit peut garder ces fonctions même si la personne effectuant le processus n'est pas l'utilisateur tracé. Néanmoins, le degré d'engagement dans le processus peut diminuer puisque il ne s'agit plus de redocumenter son activité, et des questions d'éthique et de vie privée se posent (O'Hara et *al.* 2009).

Par exemple, un enseignant qui utilise un système d'apprentissage peut redocumenter la trace de son activité informatique pour générer un rapport pédagogique d'aide et l'envoyer à ses étudiants. Dans ce document, il peut expliquer la façon de résoudre un problème particulier en utilisant le système (actions à effectuer) ou mettre l'accent sur les ressources qui lui semble intéressantes à exploiter. De façon similaire, chacun des étudiants peut redocumenter la trace de son activité d'utilisation du système pour décrire ce qu'il a fait exactement et pourquoi, les difficultés qu'il a rencontrées, les erreurs qu'il a commises, etc. Le document produit, une fois archivé, peut servir l'étudiant à se rappeler de son expérience et à analyser celle-ci pour améliorer son apprentissage ; comme il peut être envoyé ou publié pour servir d'autres apprenants utilisant le même système ou des analystes de l'activité et du comportement des apprenants (tuteurs, développeurs, etc.). Parfois, le processus de redocumentation devient incontournable comme dans le cas d'un employé (e.g. une secrétaire, un travailleur de la connaissance, etc.) chargé de fournir régulièrement un compte rendu de son activité à son supérieur ; alors il peut exploiter le contenu des traces de son activité informatique à travers le processus de redocumentation pour produire un document cohérent résumant et justifiant son travail pour l'envoyer à son supérieur. De plus, en archivant ce type de documents, il crée une sorte de mémoire organisationnelle de son activité quotidienne intense.

La redocumentation encourage la personnalisation et le partage des expériences, ce qui soutient la créativité et maintient les relations sociales (Oulasvirta et Blom, 2008). La mise en place effective de ce processus doit vérifier certains critères :

– Etre à la fois facile à utiliser et utile pour les utilisateurs [4].

– Préserver la liberté des utilisateurs dans l'appropriation du contenu et de la forme du document produit.

– Fournir un produit de qualité (lisible, compréhensible, bien organisé et facilement échangeable).

Pour répondre à ces exigences, nous adoptons les principes suivants :

– Une approche *narrative* pour le processus : il s'agit de permettre à l'utilisateur de décrire naturellement l'activité à travers la construction d'une sorte de récit.

– Un processus *interactif* assisté par un *outil auteur* (de création de contenu) : pour préserver la liberté de l'utilisateur, faciliter son travail et diminuer son effort.

[4] Les u*tilisateurs* sont les individus qui effectuent le processus de redocumentation.

– Une approche de *modélisation sémantique* pour les traces et les documents : pour permettre la visualisation et la transformation automatique des traces, le maintien automatique de la cohérence des documents et leur traitement automatisé.

– Une forme textuelle pour le produit documentaire final : une forme familière et facilement échangeable.

3. Un état de l'art sur la (re)documentation de l'activité

3.1. Du Memex au lifelogging

Les recherches sur la documentation instrumentée de l'activité humaine médiatisée remontent à 1945 quand Bush a envisagé le Memex (Bush, 1945), un dispositif dans lequel un individu peut stocker et revoir ses documents et ses communications, fournissant ainsi un support étendu à sa mémoire. Plus d'un demi-siècle plus tard, des projets tels MyLifeBits (Gemmell et *al.*, 2006), LifeLog (IPTO, 2003) et SemanticLIFE (Admed et *al.*, 2004) ont implémenté cette vision à travers la collecte du plus grand nombre possible de données concernant la vie et les activités d'un individu. La technologie émergente du *lifelogging* permet de rendre une telle pratique accessible à tous en assurant la capture de différents aspects de la vie quotidienne dans des collections numériques riches et multimodales appelées *lifelogs* (emails, historiques de navigation, documents, photos, vidéos, données de localisation, etc.). Si ces lifelogs fournissent une première documentation de l'activité d'un individu ; leur taille souvent importante et leur nature hétérogène rendent leur utilisation problématique.

Pour surmonter le problème, certains projets ont opté pour l'annotation des données des lifelogs, tels que SemanticLIFE qui permet de les enrichir automatiquement avec des annotations sémantiques en vue d'interrogation avancée, ou Phothocopain (Tuffield et *al.*, 2006), un système d'annotation de collections de photos personnelles. Ce dernier système s'est plutôt concentré sur la génération d'annotations que l'utilisateur peut modifier ou enrichir afin de décrire les événements observables de la vie d'une personne ; les annotations produites sont des combinaisons de métadonnées générées pour décrire le contenu des photos et de métadonnées décrivant le contexte dans lequel ont été capturées ces photos (données de localisation, entrées d'agendas, emails, etc.). Or, (re)documenter l'activité humaine ne peut se réduire à annoter les données enregistrées à propos de celle-ci ; un tel processus implique notamment l'organisation et la visualisation de ces données de façon à les rendre plus *intelligibles* pour les utilisateurs destinataires. Une solution naturelle repose sur le paradigme narratif introduit par Fisher (Fisher, 1989), basé sur l'idée que les êtres humains sont des narrateurs (*storytellers*) quand il s'agit de décrire leurs expériences.

3.2. Du lifelogging au storytelling

Le *storytelling* vise la construction d'histoires pour décrire des idées, des émotions, des connaissances et des expériences personnelles à partager. Appliqué au domaine des lifelogs, il permet à un utilisateur de transformer une multitude d'observations, d'impressions et de données collectées pour former une unité cohérente ayant un sens pour lui (Frenzel et *al.*, 2004). La technologie du lifelogging, l'accessibilité des réseaux sociaux et la disponibilité d'outils de création et d'édition de contenu renouvellent cette tradition sous forme numérique (Strahovnik et Mecava, 2009). Les approches de *storytelling* numérique exploitent la modélisation des narrations générées et peuvent être automatiques ou semi-automatiques. La modélisation de la narration peut se faire selon trois axes (Tuffield et al., 2007) : la modélisation du comportement, la modélisation de l'utilisateur ou la modélisation de l'histoire. Les approches basées comportement utilisent des agents autonomes pour modéliser les personnages et leurs interactions afin d'aboutir à des narrations émergentes, mais les narrations produites sont souvent peu pertinentes. Les approches basées utilisateur emploient un modèle narratif explicite à adapter selon le profil utilisateur, ce qui limite leur puissance. Les approches basées histoire utilisent des structures narratives et des méthodes basées sur des règles, inspirées de théories narratives telle que RST.

Plusieurs travaux s'intéressent au *storytelling* appliqué aux lifelogs ; certains se limitent à des études expérimentales visant l'analyse du comportement de participants sollicités à décrire leurs activités à base de leurs lifelogs (Lindley et *al.*, 2009 ; Harper et *al.*, 2007), alors que d'autres s'intéressent à supporter automatiquement cette technique. Dans le projet MyLifeBits, une option « d'histoire interactive par requête » (Gemmell et *al.*, 2005) a été ajoutée pour permettre à l'utilisateur d'appliquer des requêtes sur le lifelog, de sélectionner graphiquement des photos du résultat puis de mettre ces photos en séquence, avec une possibilité de leur attribuer des légendes et des clips audio, pour composer une histoire en diaporama. Un travail plus spécifique (Kelliher et Davenport, 2007) a décrit une variété de méthodologies et de bonnes pratiques pour solliciter et enregistrer les témoignages personnels, ainsi que des stratégies de construction d'histoires avec des critères d'évaluation pour celles-ci. L'application développée permet à l'utilisateur de charger et d'éditer des données personnelles (photos, films, fichiers audio) et de construire des histoires par sélection graphique et arrangement de ces données selon des canevas en 2D ; des relations peuvent exister entre les données et des modèles prédéfinis d'histoires sont exploitées.

Appan *et al.* (2004) ont plutôt adopté une approche automatique plus poussée, basée sur l'environnement de conception et de présentation d'histoires filmiques à base d'agents computationnels de Brooks (Brooks, 1997) ; cet environnement utilise trois composants de *storytelling* —un modèle narratif (structure dramatique), une organisation des parties de l'histoire avec des représentations de leurs sens, et une stratégie de navigation à travers ces parties pour générer des narrations multi-points-de-vue. Ils ont utilisé un modèle de données simple (photos et texte), simplifié le modèle narratif à exploiter par l'utilisateur pour catégoriser les

données (introduction, conflit, etc.) et proposé des relations à ajouter entre celles-ci (ordre, conflit-résolution, etc.) pour que des agents programmés puissent construire des histoires (diaporamas) selon des règles particulières à chacun. Cependant, les histoires générées sont disjointes puisque la structure dramatique n'était pas adéquate aux données du lifelog et la relation de causalité entre événements était difficile à expliciter. Ils ont alors proposé l'utilisation d'une structure *d'histoire émergente* qui évolue à travers l'interaction de l'utilisateur ; l'outil de *storytelling* interactif qu'ils ont développé est basé sur un modèle d'événement caractérisant celui-ci par une temporalité, une localisation, des données du lifelog (photos, sons et texte) et des participants pour distinguer des points de vue ; les événements sont liés temporellement et structurellement pour permettre à l'utilisateur de varier ses navigations dans cet ensemble. Cela permet l'émergence d'histoires (diaporamas), composées progressivement à partir des données attribuées aux événements parcourus, selon différents points de vue. Par ailleurs, l'utilité de toutes les données du lifelog a été mise en doute, d'où la proposition de réduire ce contenu aux données les plus importantes.

Byrne et Jones (2009) ont opté pour l'approche proposée de *storytelling* basée-événement, mais ils ont critiqué l'idée de surmonter la complexité du lifelog en réduisant son contenu. Ils ont développé un moteur de génération de narrations basé-événement (Byrne et Jones, 2009b) pour lequel les données du *lifelog* doivent être structurées et alignées selon trois niveaux : artéfacts, épisodes et thèmes. Les artéfacts (photos, sms, documents, etc.) sont alignés temporellement et un traitement thématique leur est appliqué afin d'extraire un thème saillant pour chacun. L'utilisateur localise et sélectionne ensuite des épisodes (événements) pertinents pour l'histoire dans cet ensemble ; il fait également des choix de présentation pour que le système puisse générer l'histoire qui sera présentée comme un montage thématique qu'il pourra ensuite personnaliser.

3.3. Vers une approche semi-automatique de storytelling utilisant RST

Nous optons pour une approche *naturelle* (basée sur le *storytelling*) pour la redocumentation de l'activité informatique à partir de ses traces. D'une part, nous pensons que l'approche manuelle de construction d'histoires préserve la liberté de l'utilisateur à s'approprier la description de l'activité tracée et facilite l'expression des informations relatives à son contexte ; ce type d'approche est cependant souvent coûteux en temps et en effort pour les utilisateurs. D'autre part, nous pensons qu'une approche automatique de *storytelling* à partir des traces d'activité économise le temps et l'effort de l'utilisateur, mais elle est à la fois trop contraignante pour la liberté de l'utilisateur et ne peut fournir des histoires de qualité. Par conséquent, nous optons pour une approche semi-automatique.

Nous adoptons la vision de Bal (Bal, 1997) qui considère que tout modèle complet de la narration doit prendre en compte trois éléments : le *Fabula* pour décrire les objets/événements de base et leurs relations, l'histoire pour décrire l'arrangement des objets/événements pour un objectif particulier et la narration pour décrire comment cela est réalisé dans une forme particulière. Pour que la machine

puisse produire des narrations, l'idée est de lui fournir une compréhension de la narration à travers des représentations sémantiques (ontologies) pour ces différentes couches. Dans notre cas, la modélisation du *Fabula* sera assurée via la modélisation sémantique des traces d'activité ; alors que la modélisation de l'histoire de l'activité sera assurée via la modélisation sémantique du document produit par une ontologie basée-RST. Nous avons choisi RST parmi d'autres théories pour sa simplicité et la diversité de ses applications parmi lesquelles : la génération et l'analyse de textes (Taboada et Mann, 2006), la génération de documentaires vidéos (Bocconi et *al.*, 2005), l'aide à l'écriture (Silva et Henderson, 2005) ou le développement de modèles pour des applications de *storytelling* (Nakasone et Ishizuka, 2006). RST est également utilisée, ce qui sera important pour nous, pour le maintien de la cohérence de documents produits (Taboada et Mann, 2006b). Nous modéliserons la narration selon une forme textuelle (récit) pour laquelle des modèles de présentation prédéfinis pourront être proposés à l'utilisateur.

L'utilisateur interviendra directement durant le processus de redocumentation pour construire ce que nous nommons « histoire de l'activité ». Nous justifions ce choix par le fait qu'il est impossible de modéliser formellement tout ce qui peut être impliqué dans une activité. Une telle approche interactive nécessite de fournir un outil *auteur* à l'utilisateur pour minimiser son effort cognitif et faciliter son travail durant la construction d'une histoire, tout en assurant la cohérence de celle-ci. Si les outils que nous avons présentés ci-dessus se concentrent plus sur la présentation de l'aspect visuel des lifelogs (e.g. photos) ; d'autres outils *auteurs* non spécifiques au domaine du *storytelling* nous ont servi de source d'inspiration pour le développement de notre outil. Il s'agit de l'agent pédagogique basé-RST DPs (Rizzo et *al.*, 2002) qui aide les élèves à préparer des présentations multimédias à partir des explications de leurs cours, et du système Artequakt (Kim et *al.*, 2002) qui aide les utilisateurs à générer des biographies personnalisées à partir de fragments de données sur le web en utilisant des gabarits et des styles prédéfinis de présentation.

4. Un cadre général pour le processus de redocumentation en texte de l'activité

Le cadre général que nous avons proposé pour le processus de la redocumentation textuelle de l'activité informatique (Figure 2) implémente une approche narrative semi-automatique basée-modèles. Trois éléments sont à considérer dans ce cadre : l'entrée du processus (un ensemble de traces modélisées de l'activité), la sortie du processus (le document texte final, modélisé, décrivant l'activité), et les deux phases du processus. La première phase du processus (P0) supporte la transformation automatique de la trace(s) d'activité choisie par l'utilisateur (l'utilisateur tracé ou une autre personne) en un document texte initial fragmenté (D_0). La deuxième phase (P1) est interactive et assistée par un outil *auteur*. Elle permet à l'utilisateur de transformer le document initial via un ensemble d'opérations, jusqu'à ce qu'il soit satisfait du document produit (D_J), et que ce dernier soit exporté comme un document final (D_F). Nous décrivons tout d'abord nos modélisations des traces d'activité et des documents produits avant de détailler les deux phases du processus.

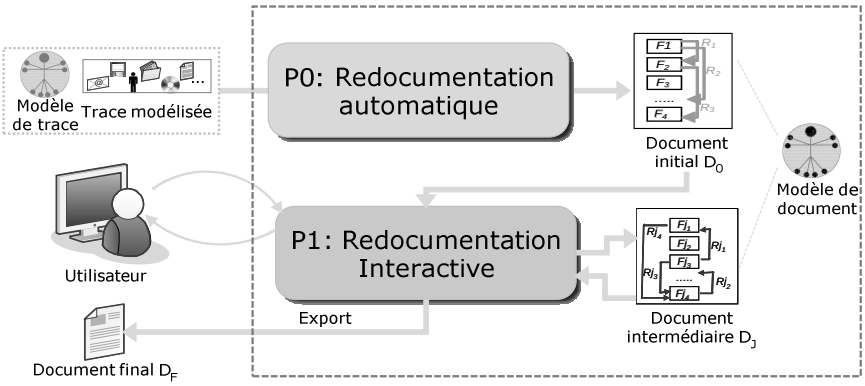


Figure 2. Un cadre général du processus de redocumentation en texte de l'activité informatique, basé-modèles (pour les traces et les documents).

4.1. La modélisation des traces d'activité

Les traces d'activités informatiques résultent de l'enregistrement des interactions des utilisateurs avec les systèmes informatiques médiatisant leurs activités. La modélisation de celles-ci dépend de l'activité tracée et de l'approche de modélisation adoptée. Plusieurs travaux ont traité cette question (Anjewierden et Efimova, 2006; Choquet et Iksal, 2007; Dragunov et *al.*, 2005) ; cependant, leurs façons de collecter, de modéliser et d'utiliser les traces d'activités différent. Notre modélisation des traces est basée sur les recherches menées au sein de l'équipe Silex [5], en particulier le méta-modèle proposé dans le cadre du Système à Base de Traces (SBT) (Laflaquière *et al.*, 2006). Le SBT est un cadre théorique et pratique pour la gestion des traces d'activités informatiques. Cela inclut la collecte, la modélisation, la transformation et la visualisation des traces. Une *trace modélisée* d'une activité médiée informatiquement est définie comme *un ensemble d'observés, temporellement situés, enregistrés sur un média, résultant de l'observation de l'utilisation d'un système informatique pour accomplir une activité particulière*. Un élément *observé (obsel)* décrit une partie de la trace d'activité ; il peut représenter une action, un objet ou une ressource manipulée, un état, un événement, etc. Quant au *modèle de la trace*, il est défini comme *un modèle formel (éventuellement une ontologie) qui décrit la structure et les propriétés des observés de la trace d'activité.*

Notre modèle de trace (Figure 3) est le produit d'un raffinement du méta-modèle de la trace d'activité du SBT en un modèle générique de trace, en considérant les actions de l'activité comme les éléments observés de sa trace. La temporalité de l'observé (action) peut être un intervalle de temps ou un instant caractérisant son occurrence. Nous supposons que la trace concerne un utilisateur particulier et qu'elle

5. Silex (équipe du laboratoire LIRIS): http://liris.cnrs.fr/silex

est composée d'un ensemble d'actions. Chaque action peut manipuler des entités (objets ou ressources). Cette considération est inspirée du modèle de l'activité unifiée (Thomas, 2005) qui représente l'activité comme une association de propriétés (ressources, produits, actions, etc.). Afin de pouvoir spécialiser le modèle générique de trace proposé pour des applications réelles, nous considérons que les actions et les entités peuvent être organisées en hiérarchies et que des relations structurelles sont possibles entre les entités (par exemple, une ressource peut composer une autre, telle la relation entre un fichier et son répertoire).

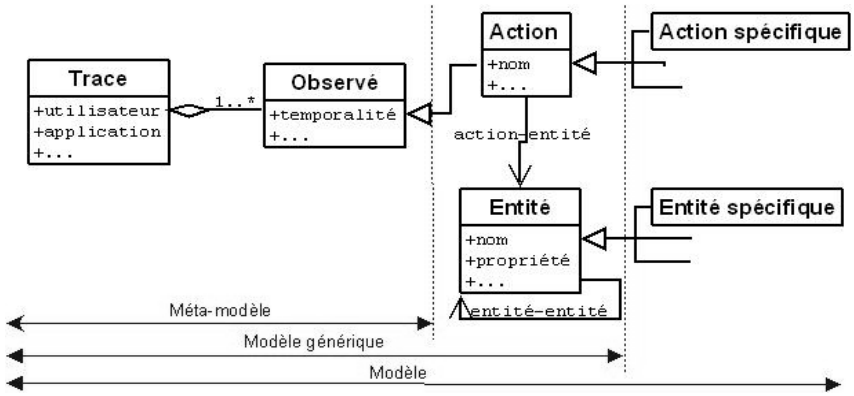


Figure 3. Modèle générique de trace d'activité (à spécialiser pour des applications).

4.2. La modélisation des documents produits

Un document numérique est une structure d'information dont le contenu peut inclure des objets simples ou composites ayant des formes différentes (texte, graphique, vidéo, audio, etc.), ou même des groupes d'objets (Martin et Alpay, 1996). Par conséquent, il est fragmenté en plusieurs unités d'information inter-liées, appelées « fragments d'information » ou « segments » (Ranwez et Crampes, 1999). Dans notre cas de redocumentation textuelle, les segments composant le document produit sont des textes inter-liés par des relations typées. Le document numérique peut être décrit selon trois niveaux différents —sémantique, logique et physique (Christophides, 1998) ; une structure particulière peut exister pour chacun des ces niveaux. La (les) structure(s) sémantique(s) organise le sens du contenu du document (e.g. la structure narrative), la structure logique décrit la structure syntaxique du document (e.g. en titres, chapitres, paragraphes, etc.) et la structure physique définit l'apparence du document et sa présentation (e.g. la typographie). Il est important de considérer toutes ces structures dans la construction d'un document. Pour cela, plusieurs études (Djemal et *al.*, 2009; Barrutieta et *al.*, 2002; Garlatti et Iksal, 2003) se sont intéressées à modéliser la multi-structuralité du document en considérant le segment comme l'unité de base sur laquelle reposent les structures.

Nous nous intéressons particulièrement au contenu du document texte produit à travers le processus de redocumentation et au maintien de sa cohérence lorsque celui-ci est manipulé par l'utilisateur. Nous focalisons notre modélisation du document sur sa structure sémantique, sans pour autant négliger les deux autres structures (logique et physique) ; nous admettons que l'élaboration de n'importe quel document implique la définition d'une application entre sa structure sémantique et la structure du document final tel qu'il sera présenté (Nanard, 1989). Notre modélisation (Figure 4) est faite selon deux niveaux d'abstraction : un modèle générique (partie supérieure) qui peut être spécialisé en un modèle spécifique du document (partie inférieure). Pour le modèle générique, un document est composé de plusieurs structures typées, chaque structure est composée de segments et de relations, et chaque relation peut avoir un ou plusieurs segments sources ou cibles. Quant au modèle spécifique du document, trois structures de base sont considérées (sémantique, logique et physique). La structure sémantique relie un ensemble de segments de contenu par des relations sémantiques ; alors que le type de chaque relation est déduit du type de sa structure sémantique.

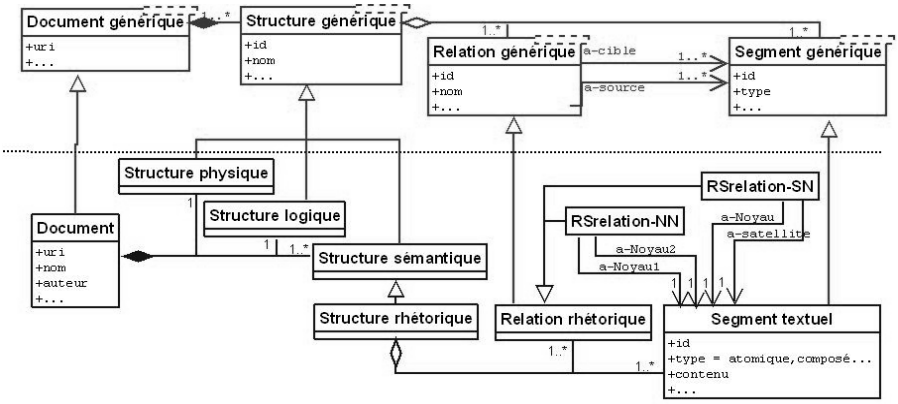


Figure 4. Modèle de document comme une spécialisation d'un modèle générique et centré sur la structure rhétorique.

La structure rhétorique est la structure sémantique du document sur laquelle nous focalisons notre modélisation. Elle est définie en termes d'éléments décrits par RST. Cette théorie a été proposée initialement pour les documents textes, ce qui convient au mieux à nos besoins. Elle permet de diviser un document texte en segments (de type Noyau ou Satellite), avec des relations rhétoriques entre ces segments dont les effets combinés assurent la cohérence de l'ensemble. Les relations rhétoriques sont souvent utilisées pour développer une argumentation liant ensemble des éléments d'un contenu. Cela inclut la cause et la conséquence, l'interprétation et la justification, la jonction et le contraste, etc. Elles sont classées comme relations *Satellite à Noyau* (RSrelation-SN), ou relations *Noyau à Noyau* (RSrelation-NN) (Taboada et Mann, 2006b). Dans le premier cas, un segment (le Noyau) est

plus essentiel pour le contenu du document que l'autre, comme par exemple un segment texte décrivant une action d'une activité qui est justifiée (relation *Justify*) par un autre segment. Dans le deuxième cas, les deux segments ont la même importance, comme par exemple pour deux segments textuels mis en séquence (relation *Sequence*) dans un document pour décrire deux actions consécutives d'une activité.

La structure rhétorique du document peut transmettre le sens de son contenu, du fait qu'elle attribue une fonction particulière à chacun des segments de ce contenu. En outre, cette structure peut être facilement implémentée comme un modèle formel (une ontologie) ; ce qui facilitera la création de documents cohérents en se basant sur des règles rhétoriques à implémenter et à respecter. La structure logique du document est composée à partir des segments et des relations de la structure rhétorique (e.g. les segments de la structure rhétorique reliés par des relations de type *Séquence* donneront dans la structure logique une suite de paragraphes reprenant leurs contenus textuels). Similairement, un modèle prédéfini de présentation peut être utilisé pour la structure physique du document (e.g. une feuille de style), avec une possibilité d'intervention de l'utilisateur pour le choix de certaines propriétés du modèle (couleur, font, etc.).

4.3. Les phases du processus de redocumentation

4.3.1. La phase de redocumentation automatique

Cette phase représente la partie automatique du processus de redocumentation et permet de transformer une trace d'activée en entrée en un document texte fragmenté initial (D_0). Ceci est rendu pour nous possible par la modélisation formelle des traces et la disponibilité d'outils dédiés à la génération de texte à partir de données modélisées. Dans le document initial D_0, chaque segment ou fragment généré (Fi) est une unité d'information décrivant un ou plusieurs observés (*obsels*) de la trace d'activité ; il est caractérisé par un contenu textuel et d'autres propriétés (e.g. l'id de l'*obsel* décrit pour garder des liens avec la trace). Quant aux relations (Ri) générées automatiquement entre les segments, elles peuvent être de type structurelle ou rhétorique. Les relations structurelles décrivent celles qui existent explicitement entre les éléments de la trace d'activité (e.g. une relation de composition entre deux objets manipulés par deux actions dans la trace d'activité) ; alors que les relations rhétoriques expriment des relations déduites de la trace (e.,g., les relations temporelles entre les *actions* de la trace peuvent être exprimées par des relations rhétoriques de type *Sequence* [6] entre les segments textuels décrivant ces *obsels* dans le document initial D_0). La phase de redocumentation automatique peut être paramétrée pour permettre à l'utilisateur de ne choisir qu'une partie de la trace à redocumenter (e.g. des types d'actions à considérer pour que seuls les segments

6. Nous considérons que la relation d'ordre temporel entre les actions de la trace est une propriété importante à exprimer dans le produit de la redocumentation ; alors pour l'exprimer sous une forme textuelle, nous avons utilisé une relation rhétorique adaptée (*Séquence).*

décrivant les actions ayant ces types soient générés, un intervalle de temps pour que seules les actions effectuées durant cet intervalle soient décrites dans D_0, *etc*.).

4.3.2. La phase de redocumentation interactive

Durant cette phase, l'utilisateur est assisté par un outil *auteur* et la génération du document final à partir du document initial (D_0) est à la fois interactive et itérative. Il s'agit d'une composition personnalisée des trois niveaux sémantique, logique et physique du document, exploitant les segments et les relations définis au sein du document initial. L'utilisateur est assisté pour effectuer des opérations sur le document (D_0) ou sur un document intermédiaire (Dj) et ainsi modifier la structure sémantique de celui-ci. Ces opérations peuvent être appliquées à un ou à plusieurs segments, à une relation entre deux segments, ou au document en entier. Elles peuvent être organisationnelles (e.g. supprimer une relation, fusionner des fragments, etc.) ou rhétoriques (e.g. introduire le document, justifier un segment, etc.). Chaque opération rhétorique correspond à la création d'une relation rhétorique particulière et d'un segment textuel satellite (dont le contenu est renseigné par l'utilisateur) qui sera lié au document ou à l'un de ses segments par cette relation.

Tableau 1. Relations rhétoriques à utiliser pour la redocumentation interactive.

RSrelation N-N	**Segment Noyau**	**Segment Noyau**
Séquence (Sequence)	texte	texte suivant
Contraste (Contrast)	description d'une possibilité	description de l'alternatif
Jonction (Joint)	texte	texte à joindre
RSrelation N-N	**Segment Noyau**	**Segment Satellite**
Justification (Justify)	texte	information légitimant l'énonciation
Élaboration (Elaboration)	information de base	information supplémentaire
Arrière-plan (Background)	texte	texte facilitant la compréhension
Facilitation (Enablement)	description d'une action	information aidant à accomplir l'action
Évaluation (Evaluation)	description d'une situation	commentaire évaluant la situation
Interprétation(Interpretation)	description d'une situation	interprétation de la situation
Introduction (introduction)	texte	introduction au texte
But (Purpose)	description d'une situation	description de l'intention
Résumé (Summary)	texte	résumé du texte
Solution (Solutionhood)	description d'une solution	description d'un problème (besoin)

Le tableau 1 présente un sous ensemble des relations rhétoriques que nous jugeons pertinentes pour la redocumentation d'une activité médiée informatiquement. Nous justifions ce choix par un compromis entre le pouvoir descriptif de l'activité informatique et la facilité d'exploiter cet ensemble de relations par l'utilisateur effectuant la redocumentation.

Comme résultat, chaque fragment textuel dans le document intermédiaire (D_J) peut décrire : (i) une action de la trace d'activité, (ii) un segment textuel composé à partir des segments du document produit à l'itération précédente (D_{J-1}), ou (iii) un segment textuel ajouté par l'utilisateur. Quant aux relations entre les segments de (D_J), elles décrivent les relations existantes dans le document (D_{J-1}) ou celles ajoutées explicitement par l'utilisateur (rhétoriques ou organisationnelles). Une fois que l'utilisateur est satisfait du document intermédiaire, il pourra l'exporter comme un document texte final. Pour cela, à chaque relation rhétorique existante dans le document (D_J) va correspondre son expression textuelle dans le document final ; une expression qui peut être explicitement choisie par l'utilisateur durant l'application des opérations rhétoriques sur le document intermédiaire. Pour la présentation du document finale, un modèle prédéfini est utilisé, mais avec la possibilité de modifier certaines de ses propriétés par l'utilisateur (e.g. la typographie et les couleurs). Nous présentons des exemples concrets de l'application des différentes opérations décrites ci-dessus dans la section consacrée à la description de notre outil de redocumentation ActRedoc.

5. Exemple d'une activité à redocumenter : navigation dans la « Presse illustrée»

Nous avons choisi une activité informatique à tracer et à redocumenter afin de mieux présenter l'implémentation du cadre général que nous avons proposé pour le processus de redocumentation. Il s'agit d'une activité de navigation dans un site d'archive de journaux du 19ème siècle de la presse illustrée régionale, développée à la bibliothèque municipale de Lyon (Figure 5). Chaque journal dans ce corpus est composé de fascicules, le fascicule est composé de pages, et la page comporte des articles et/ou des illustrations. Tous ces éléments sont décrits par des métadonnées pour supporter une recherche sémantique (par type, par mot-clé, par thème, *etc*.).

5.1. Le fichier log produit du traçage de l'activité de navigation

Le traçage de l'application nous a permis de récupérer un fichier log décrivant l'activité de navigation d'un ensemble d'utilisateurs que nous avons traité. Dans ce fichier, chaque ensemble de lignes correspond à la trace (de base) d'activité d'un utilisateur ; alors que les colonnes décrivent les informations suivantes : la date et l'heure de l'action effectuée, l'adresse IP de l'utilisateur, l'identifiant de l'action ou l'url de l'entité affichée (article, page, liste de résultats, etc.), le type de l'entité manipulée et les propriétés de celle-ci. La figure 6 représente une partie de ce fichier log. Nous pouvons déduire par exemple que l'utilisateur ayant l'adresse IP

65.55.210.151 a affiché la page n°5 du fascicule n°101 du journal Le Progrès Illustré (http://collections.bm-lyon.fr/presseIllustree/PER0012495) le 16/07/2008 à 12:30.

Figure 1. Interface de navigation dans la Presse illustrée (http://collections.bm-lyon.fr/presseIllustree/).

Temps	Adresse IP	URL / Action	Type	Propriétés
16/07/2008 12:30	65.55.210.151	/PER0012495	page	Le Progrès illustré, n°101, pp.5
16/07/2008 12:30	81.247.174.25	/PER0011787	page	Le Progrès illustré, n°37, pp.2
16/07/2008 12:31	81.247.174.25	/PER00110015	item/ illustration	La famille impériale 1891/08/30
16/07/2008 12:31	65.55.210.151	/PER0013376	page	Le Progrès illustré, n°190, pp.6
16/07/2008 12:31	81.247.174.25	<pleine page>		

Figure 6. Un extrait du log produit du traçage de l'application « Presse illustrée ».

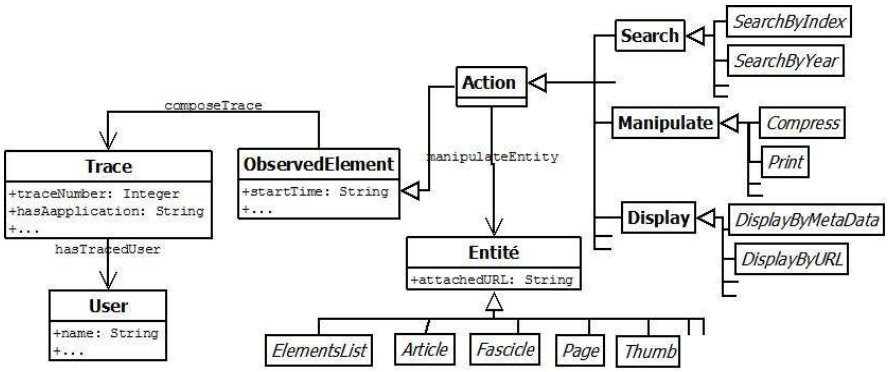


Figure 7. Extrait de l'ontologie de la trace de navigation dans la « Presse illustrée ».

5.2. La modélisation des traces de l'activité de navigation

Nous avons développé sous la forme d'une ontologie OWL le modèle de la trace d'activité de navigation dans la « Presse illustrée » comme une spécialisation du modèle générique de trace que nous avons proposé (voir section 4.1). Nous avons spécialisé les actions (e.g. display, manipulate, etc.) et les entités (article, illustration, etc.) comme c'est illustré par la figure 7 qui présente un extrait de ce modèle de trace. Les traces modélisées de l'activité sont générées à partir du fichier log récupéré selon le modèle de trace décrit par un processus automatisé, dit de collecte.

6. ActRedoc : un outil de redocumentation en texte

L'outil de redocumentation ActRedoc que nous avons développé pour la redocumentation en texte de l'activité informatique est un outil *auteur*. Il assure, à la fois, une présentation intelligible des traces d'activité facilitant ainsi leur manipulation, et une assistance à l'utilisateur pour définir le contenu et la forme du document produit et maintenir sa cohérence. L'architecture de cet outil (Figure 8), implémenté comme une application Java, est centrée autour de deux modules : le module de transformation automatique et le module de transformation interactive. Chacun de ces deux modules correspond à une phase de redocumentation de notre cadre général et est accessible à travers l'interface graphique de l'outil.

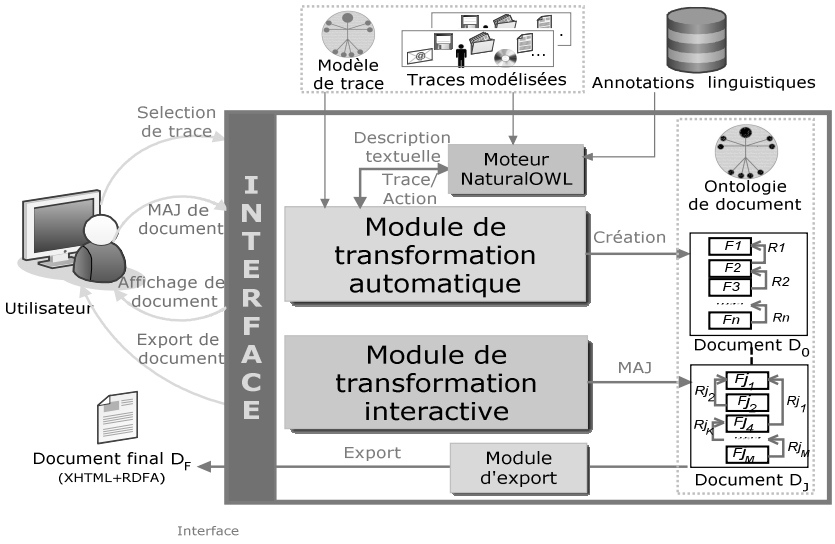


Figure 8. Architecture de l'outil de redocumentation en texte ActRedoc.

L'entrée de l'outil ActRedoc est un ensemble de traces d'activité modélisées sémantiquement par l'ontologie de la trace d'activité (en général). Pour pouvoir automatiser en partie le processus de redocumentation de l'activité, nous avons utilisé l'outil de génération du langage naturel NaturalOWL (Galanis et Androutsopoulos, 2007) ; un outil capable de générer des descriptions textuelles (en Anglais) à partir des éléments d'une ontologie en OWL (ontologie de la trace dans notre cas). Cependant, avant de pouvoir utiliser cet outil, une annotation linguistique préalable de l'ontologie par cet outil est exigée. La sortie de l'outil ActRedoc est un document texte, modélisé sémantiquement par une ontologie de document, décrivant l'activité tracée et pouvant être exporté dans une forme hypertexte enrichie avec des métadonnées par le biais du module d'export. Nous décrivons ci-dessous chacun des composants de l'outil ActRedoc tout en présentant un exemple de la redocumentation de l'activité de navigation dans la « Presse illustrée ».

6.1. Le module de transformation automatique

A travers l'interface de l'outil ActRedoc, l'utilisateur peut choisir graphiquement une trace particulière (une instance) de l'activité informatique à redocumenter. Le module de transformation automatique extrait, pour cette trace, tous les observés de type *Action* et les ordonne temporellement. Pour la trace choisie et pour chacune de ses actions (observés), il invoque le moteur NaturalOWL pour générer une description textuelle. Ce moteur exploite alors l'ontologie de la trace et les fichiers des annotations linguistiques de celle-ci pour pouvoir produire de telles descriptions. Le module de transformation automatique améliore par la suite les descriptions textuelles produites par NaturalOWL en appliquant automatiquement un ensemble de règles linguistiques de réécriture (e.g. éviter la répétition d'une même date dans la description de deux actions consécutives de la trace). Les descriptions textuelles résultantes sont utilisées dans la composition d'un document texte fragmenté D_0.

6.1.1. Les annotations linguistiques de l'ontologie de la trace

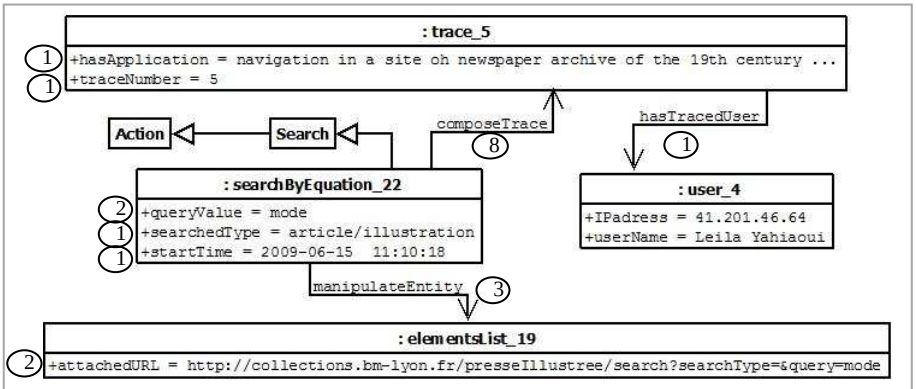
Nous avons annoté, au préalable, l'ontologie de la trace d'activité (de navigation dans la Presse illustrée) par le module d'annotation de Natural-owl [7]. Cela a permis de produire trois fichiers de métadonnées (en RDF) qui décrivent trois niveaux différents: le niveau lexical de l'ontologie, le niveau micro-plans de l'ontologie, et le niveau préférences concernant la génération des descriptions textuelles par NaturalOWL. Ce sont ces métadonnées qui permettent à NaturalOWL de générer des descriptions textuelles à partir des instances de l'ontologie annotée.

Au niveau lexical, l'annotation linguistique consiste à exprimer le genre et le nombre des classes et des instances de l'ontologie (e.g. pour la classe *SearchByEquation*, son expression au singulier est « *search action by a complex query* », au pluriel c'est « search actions by complex queries », elle est de genre neutre et comptable). Au niveau micro-plans, l'annotation linguistique concerne

[7]. Le module d'annotation de NaturalOWL s'intègre comme un plugin dans notre éditeur d'ontologies Protégé-owl (http://protege.stanford.edu/overview/protege-owl.html)

l'expression des propriétés des classes et des relations entre celles-ci selon des *templates* (e.g pour la relation *composeTrace* qui existe entre *Action* et *Trace*, nous l'avons exprimé selon le *template* suivant : propriétaire de la relation (un pronom, nominatif) + verbe (au singulier : belongs, au pluriel : belong, temps: présent, voix: active) + préposition (to) + chaine de caractères (a trace)). De plus, à chacune de ces propriétés et relations, nous avons assigné un ordre de priorité qui détermine l'ordre de son apparition dans la description textuelle générée pour une instance ayant ces propriétés et relations au même temps (e.g. la valeur 1 désigne une priorité supérieure pour *startTime*). Au niveau préférences, l'annotation linguistique consiste à préciser des paramètres qui influencent directement le processus de génération des descriptions textuelles du moteur NaturalOWL tels que le langage, la profondeur des inférences dans la description des instances (e.g. la valeur 2 pour s'arrêter dans la description textuelle d'une instance à la description des instances directement liées à celle-ci), le nombre maximum de propriétés et de relations décrite par phrase (nb-faits/phrase), *etc*.

La figure 9 représente l'exemple d'une description textuelle générée par NaturalOWL pour une instance de type Action (searchByEquation_22) en exploitant les annotations linguistiques de l'ontologie de la trace d'activité avec les préférences (langage = anglais, profondeur d'inférence = 2 et nb-faits/page = 4).



Texte généré par NaturalOWL (pour l'instance searchyEquation_22)

This is a search action by a complex query, a kind of search action, for type "article/illustration". At 2009-06-15T11:10:18 it appears, its query expression value is "mode" and it manipulates an entity, a list of results attached to the URL:http://collections.bm-lyon.fr/presseIllustree/search?searchType=&query=mode. It belongs to a trace. This trace's number is 5, it is about the application of navigation in a site of newspaper archive of the nineteenth century (CANU19 project with the municipal library of Lyon- France) and it concerns a user.

Figure 9. Un exemple de texte généré par NaturalOWL pour une instance de type Action d'une trace d'activité de navigation dans la « Presse illustrée » .

4.3.2. La composition du document initial (D_0)

Le document produit par le module de transformation automatique est constituée d'un ensemble de segments textuels dont les contenus sont générés par le moteur NaturalOWL. Ces segments sont liés par des relations rhétoriques de type *Séquence*, générés automatiquement par le module de transformation automatique pour décrire principalement la temporalité de la trace d'activité à redocumenter. La figure 10 est une représentation graphique du document (D_0) généré à partir d'une trace (trace_5) de l'activité de navigation dans la « Presse illustrée ».

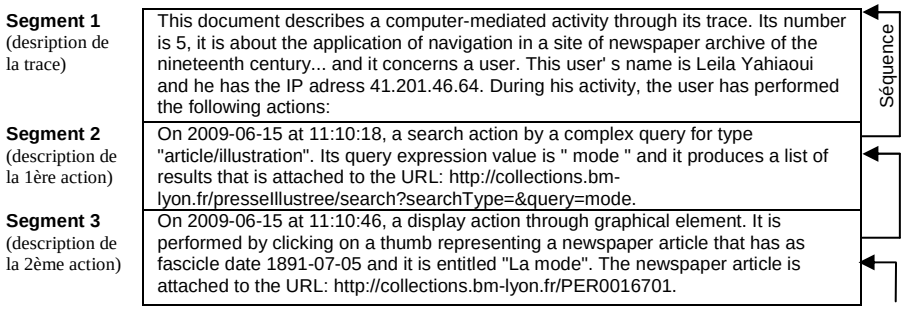


Figure 9. Une représentation graphique d'un exemple de document D0 produit par le module de redocumentation automatique (pour trace-5).

Le contenu du document (D_0) est modélisé par l'ontologie de document ; une spécialisation du modèle générique de document que nous avons proposé (voir section 4.2). Elle est inspirée en partie de l'ontologie ontoRest (Naja-Jazzar et al., 2009) qui modélise les relations de RST dans un document texte construit en collaboration et permet de détecter automatiquement les incohérences qu'il puisse contenir. La figure 11 en propose une représentation graphique de cette ontologie dans laquelle chaque document est identifié et lié à la trace qu'il décrit (propriété *uriTrace*). Il a un segment particulier, sa racine, auquel les autres segments du document sont liés. Chaque segment appartient à un seul document et est caractérisé par des propriétés. La propriété *Uri-source* réfère l'instance *trace* ou *action* décrite par le segment ; alors que *typeSegment* indique si le segment est original (issu de la trace), composé (par fusion de segments), ajouté ou remplacé par l'utilisateur. Chaque segment a un contenu textuel et une position (d'ordre) dans le document qu'il compose. Quant aux relations rhétoriques entre les segments, elles sont de deux types (*RSrelation-SN* et *RSrelation-NN*) et sont décrites dans le tableau 1.

Pour composer le document (D_0), le module de transformation automatique commence par créer une instance *document* dans l'ontologie de document. Le contenu du premier segment du document (racine) est la description textuelle générée à propos de la trace d'activité, enrichie avec une expression textuelle ajoutée automatiquement pour introduire le reste du document. Le contenu du deuxième segment créé est la description textuelle générée à propos de la première

action de la trace d'activité ; ce segment est relié au segment racine par une relation rhétorique de type *Séquence*. De la même façon, les autres segments du document sont créés à partir du reste des actions de la trace ; ils sont ordonnés et liés dans le document (D_0) selon l'ordre temporel des actions correspondantes dans la trace. Ce document initial fournit une première description de l'activité tracée et peut être exporté en texte brut ou en hypertexte, facilitant ainsi sa présentation et son échange. De plus, il sert d'entrée pour la phase de redocumentation interactive.

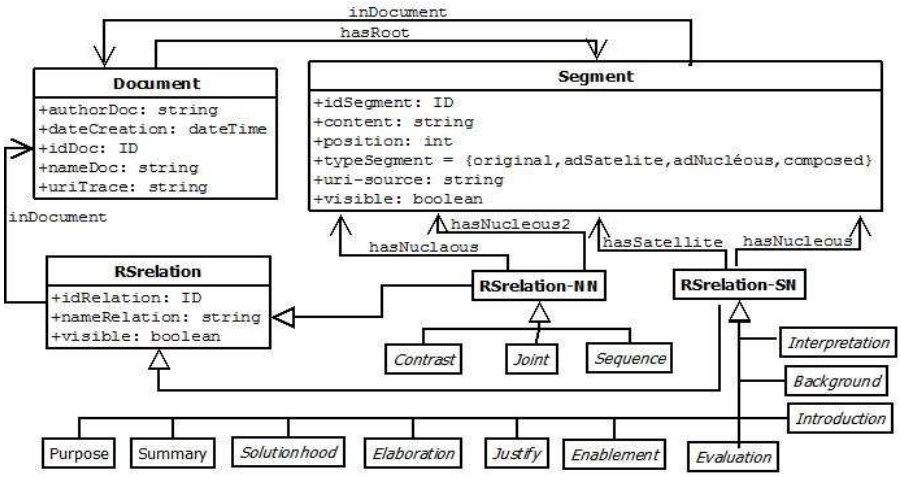


Figure 11. Représentation graphique de l'ontologie de document.

6.2. Le module de transformation interactive

Le module de transformation interactive permet à l'utilisateur d'appliquer itérativement des opérations de transformation sur le document (D_0) et de générer un document intermédiaire (Dj) après chaque itération, jusqu'à ce que l'utilisateur soit satisfait du document produit alors appelé document final. Ces opérations, que nous résumons dans le tableau 2, peuvent être rhétoriques ou organisationnelles. Elles sont accessibles via l'interface de l'outil ActRedoc (Figure 12).

Cette interface est composée de trois panneaux. Le panneau de droite permet l'accès aux opérations de transformation. Le panneau central affiche les segments textuels du document manipulé (D_j) ; la couleur de contour de chaque segment renseigne sur son type (e.g noir pour un segment original, bleu pour un segment composé, etc.) et une bulle contextuelle d'informations permet d'afficher ses autres propriétés (sa position et ses relations rhétoriques avec d'autres segments). Le panneau de gauche affiche le document intermédiaire dans une forme hypertexte en utilisant une feuille de style prédéfinie ; des métadonnées (titre, auteur, etc.) sont saisies par l'utilisateur et assignées au document au moment de sa présentation.

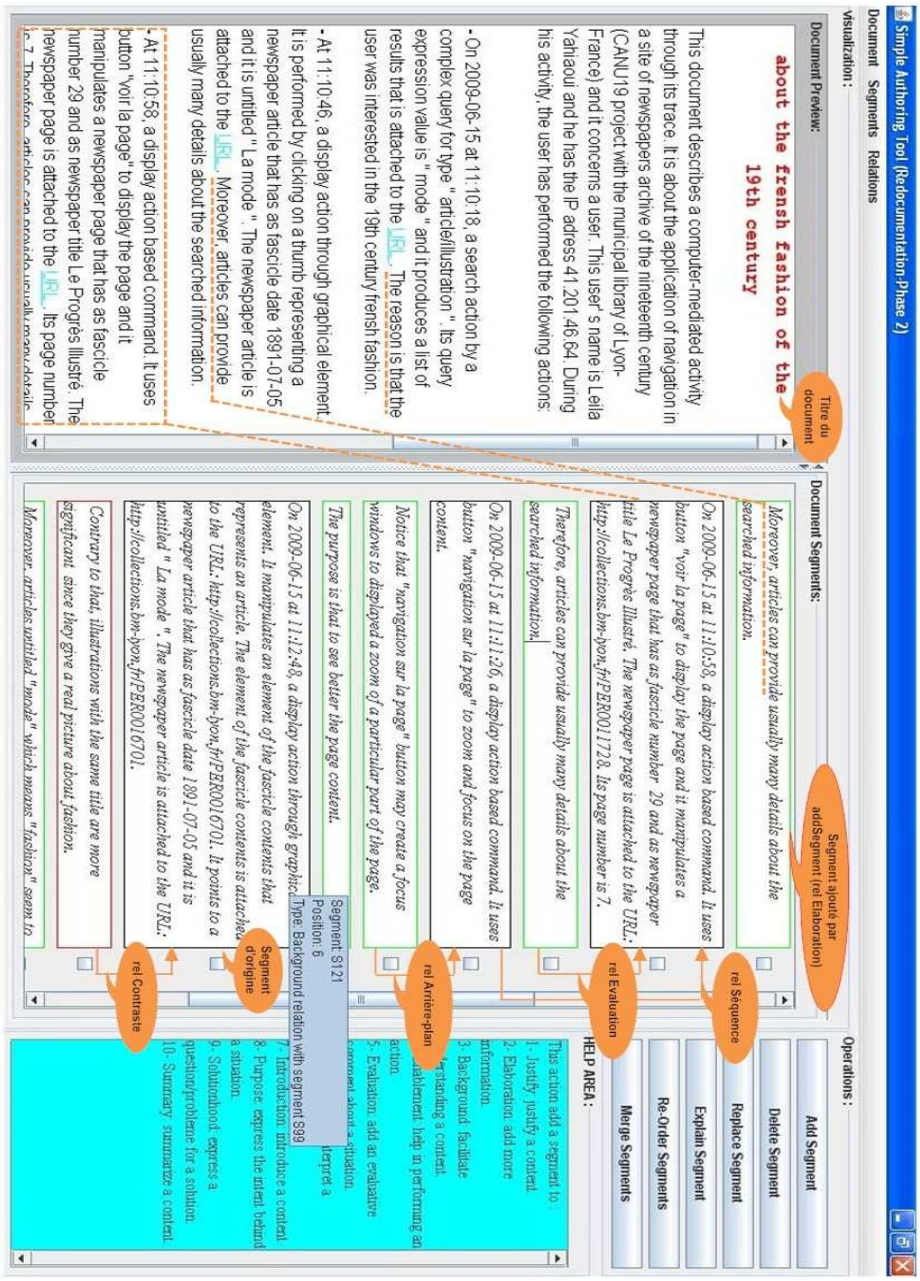


Figure 12. Interface de l'outil de redocumentation en texte ActRedoc (avec l'exemple d'un document produit de trace_5) .

Tableau 2. Ensemble des opérations de transformation supportées par ActRedoc.

Opération	Description
add segment (rhétorique)	Ajouter un segment au document avec une relation N-N Le contenu du nouveau segment est à saisir par l'utilisateur et le nouveau segment sera lié au segment (du document) choisi par l'utilisateur avec une relation rhétorique de type N-N telle que Joint (voir tableau1).
delete Segment (organisationnelle)	Supprimer un segment du document Pour les relations qui existent entre ce segment et d'autres segments, celles de type Séquence seront mises à jour et les autres supprimées.
replace Segment (organisationnelle)	Remplacer le contenu d'un segment existant avec un nouveau texte Le nouveau texte est à saisir par l'utilisateur.
explain segment (rhétorique)	Ajouter un segment au document avec une relation S-N Le contenu du nouveau segment est à saisir par l'utilisateur et le nouveau segment sera lié au segment (du document) choisi par l'utilisateur avec une relation rhétorique de type S-N telle que Justify (voir tableau1).
reorder segments (organisationnelle)	Déplacer un segment du document par rapport à un autre segment (avant ou après) ; mettre à jour les relation de type Séquence concernées
merge segments (organisationnelle)	Fusionner les contenus de segments consécutifs dans un document Les relations de type Séquence entre ces segments seront supprimées et les relations de ces segments avec d'autres segments seront mises à jour.

Dans l'exemple de document affiché, les URLs surlignées sont des liens vers des ressources manipulées par des actions de la trace d'activité ; alors que les relations rhétoriques de types *Séquence* entre les segments du document sont représentées par le positionnement séquentiel des contenus textuels de ces segments dans le document affiché. Quant aux autres types de relations rhétoriques contenues dans le document, elles sont exprimées en texte selon les choix de l'utilisateur. Le tableau 3 présente les expressions des relations rhétoriques proposées à l'utilisateur au moment de la création de ces relations. Par exemple, si l'utilisateur applique l'opération *Explain segment* sur un segment du document et choisit *Elaboration* comme type de relation rhétorique à ajouter, l'outil ActRedoc lui propose de faire un choix dans une liste d'expressions textuelles correspondantes à cette relation (e.g « *Moreover* »). Le but est de pouvoir exprimer automatiquement cette relation rhétorique par du texte au moment de la présentation du document produit (*preview*). Ainsi, l'expression choisie sera placée entre le contenu textuel du segment cible de la relation (segment décrivant l'idée de base) et celui du segment ajouté (segment fournissant une élaboration de cette idée). Les lignes en pointillés sur la figure 12 montrent des correspondances entre des segments (et des relations) du document manipulé (partie centrale de l'interface) et des parties de sa présentation (partie gauche de l'interface).

Tableau 3. Expression des relations rhétoriques pour la présentation du document.

Opération	Description
Séquence	Texte du segment cible (de la relation) → texte du segment source.
Contraste	Texte du segment cible → « Contrary to that » → texte du segment source.
Jonction	Texte du segment cible → « And » → texte du segment source.
Justification	Texte du segment cible → «The reason is that » ou « that is because » → texte du segment source.
Élaboration	Texte du segment cible → « Moreover » → texte du segment source.
Arrière-plan	Texte du segment cible → « Notice that » → texte du segment source.
Facilitation	Texte du segment cible → « To do that » → texte du segment source.
Évaluation	Texte du segment cible → « Therefore » ou « So » → texte du segment source.
Interprétation	Texte du segment cible → « This means that » → texte du segment source.
Introduction	« Introduction: » → texte du segment source → texte du segment cible (ou le contenu du document si c'est tout le document qui est introduit).
But	Texte du segment cible → « The purpose is that » → texte du segment source.
Résumé	Texte du segment cible (ou le contenu du document si c'est tout le document qui est résumé) → « Summary: » → texte du segment source.
Solution	Texte du segment cible→ «This was a solution for the case»→ texte segment source.

6.2. Le module d'export du document final

Une fois que l'utilisateur est satisfait du document intermédiaire (Dj), ce document peut être exporté par le module d'export dans une forme hypertexte enrichie avec des métadonnées. En plus des métadonnées ajoutées explicitement par l'utilisateur (titre, date et auteur) et qui sont incorporées automatiquement dans l'entête du document ; des métadonnées sont ajoutées automatiquement au document pour décrire son contenu et sa structure en utilisant RDFa (Adida et Birbeck, 2008). Notre objectif est de permettre l'annotation des données exploitables par les humains (texte du document produit) avec des indicateurs exploitables par les machines (faisant référence aux éléments de l'ontologie de trace et de celle de document) pour les navigateurs et d'autres programmes d'interprétation. La figure 13 montre une représentation du contenu sémantique du document produit. Le

module d'export utilise un gabarit prédéfini pour la structure logique et une feuille de style pour la structure physique du document, mais ces modèles peuvent être paramétrés pour supporter les choix de l'utilisateur dans la mise en forme du document final.

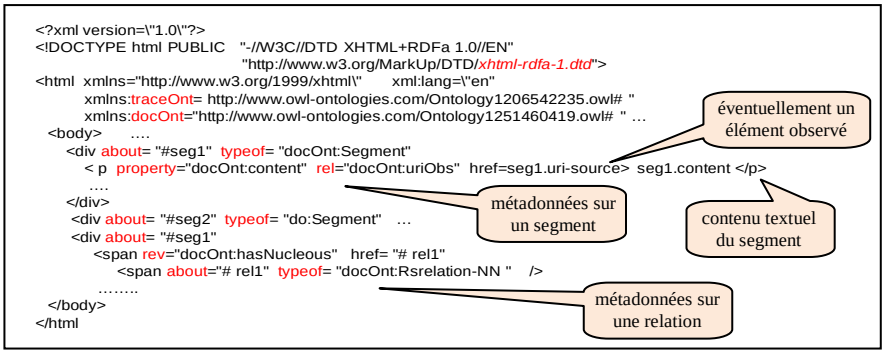


Figure 13. Vue du contenu du document final exporté (XHTML+RDFa).

7. Une première évaluation de l'outil de redocumentation ActRedoc

Nous avons testé notre prototype ActRedoc dans le cadre d'une étude préliminaire visant l'évaluation de son utilisabilité [8] et de son utilité [9]. Une telle évaluation nécessite souvent un nombre important d'utilisateurs afin de surmonter les variations de la performance entre individus, ce qui risque de rendre cette évaluation couteuse et parfois trompeuse. Cependant, une étude de Nielsen a montré qu'un nombre restreint d'utilisateurs (typiquement 5) est suffisant pour une première évaluation d'un système (Nielsen et Loranger, 2006) ; la révision de la conception étant suivie d'autres évaluations similaires afin d'améliorer le système de façon progressive et peu couteuse. C'est l'idée que nous avons adoptée dans l'évaluation de notre outil de redocumentation. L'activité tracée étant celle de navigation dans la Presse Illustrée (voir section 5), un ensemble d'utilisateurs a été sollicité pour appliquer des scénarios de navigation pertinents ; les traces de leurs activités (fichier log) ont été modélisées sémantiquement selon l'ontologie de trace (voir Figure 7). Afin de montrer à travers un ensemble minimal de traces d'activité (*i*) l'effet de la longueur des traces sur le processus de redocumentation, et (*ii*) l'effet de la durée de temps entre le traçage de l'activité et sa redocumentation sur le processus, nous avons sélectionné dans l'ensemble seulement six traces ayant les caractéristiques suivantes : (*i*) des traces de longueurs différentes (en nombre d'actions) dont quatre traces ont une longueur similaire, une trace de longueur maximale (T1) et une autre

[8]. L'utilisabilté d'un outil est définie en termes de mesure de la facilité de son utilisation: temps d'apprentissage, efficacité d'utilisation, satisfaction des utilisateurs, *etc*.
[9]. L'utilité d'un outil réfère sa fonctionnalité et s'il répond aux besoins des utilisateurs.

de longueur minimale (T5), et (*ii*) deux traces concernent un même utilisateur, celle d'une activité ancienne (T1) et celle d'une activité récente (T2).

7.1. Expérimentation

Nous avons demandé à dix participants de redocumenter les traces d'activité choisies en utilisant notre outil de redocumentation. Parmi ceux-ci, cinq participants ont déjà effectué l'activité informatique tracée (utilisateurs tracés) alors nous avons demandé au reste de consulter rapidement l'application tracée. L'attribution des traces aux participants n'était pas faite de façon purement arbitraire puisque chacun des utilisateurs tracés devait redocumenter sa propre activité et un participant (P1) devait redocumenter à la fois une activité ancienne et une autre récente. De plus, chacune des traces d'activité devait être exploitée par au moins deux participants et un même participant (P8) devait redocumenter la trace la plus longue et celle la plus courte. Nous avons demandé aux participants de décrire oralement les activités qu'ils devaient redocumenter à partir de leurs traces (fichier log). A ce niveau, nous avons constaté que ces participants avaient des difficultés à interpréter le contenu des traces, en particulier lorsque la trace d'activité n'était pas celle du participant. Nous avons initié les participants à l'utilisation de notre outil ActRedoc à travers un exemple de redocumentation, en utilisant les différentes opérations supportées par l'outil. Les participants ont procédé, par la suite, à des tests séparés de l'outil pour redocumenter les traces qui leurs ont été attribuées, sans pour autant leur imposer des contraintes à propos de l'objectif du processus de redocumentation. Durant ces tests, nous avons observé chacun des participants et nous avons pris des notes à propos de ses attitudes (difficultés rencontrées, utilisation des opérations, durée, *etc*).

7.2. Résultats

Une fois l'expérimentation achevée, nous avons demandé à chacun des participants de décrire l'objectif de sa redocumentation et d'évaluer l'outil utilisé. Ces interviews ont révélé une facilité de l'utilisation de l'outil par les participants, ainsi que leur satisfaction à propos de l'ergonomie de son interface, de ses fonctionnalités et de ses produits. Nous avons résumé nos observations et les réponses des participants dans le tableau 4 comme suit :

– La première colonne affiche le numéro du participant, le numéro de la trace qu'il a redocumenté et le numéro de l'utilisateur tracé.

– La deuxième colonne affiche la durée du processus de redocumentation (mn).

– La troisième colonne décrit le genre du document produit que nous qualifions de rapport d'analyse (R-ana), de rapport de synthèse (R-syn), ou de rapport annoté (R-ano) ; il peut être personnel (P) ou destiné à d'autres personnes (A) et long (L) ou court (C). Cette classification est basée sur notre analyse des contenus et des formes des documents produits ainsi que les objectifs des participants. Dans un rapport de synthèse, le participant extrait du document initial (D_0) décrivant la trace les informations qui lui semblent importantes, puis s'approprie ce contenu selon

Tableau 4. Synthèse de l'expérimentation de redocumentation et de ses résultats.

Participant, Trace, utilisateur tracé	**Durée du processus**	**Document produit** (genre, destina, taille)	**Utilité de l'outil**		**Utilisabilité**				
			Evaluation de D_0	**Opérations** utilisées (U) / non utilisées (NU) / fréquemment utilisées (F)	**Evaluation besoin**	**Apprenabilité**	**Efficacité**	**Mémorabilité**	**Satisfaction**
1,1,1	40	R-ano, P, L	B	NU : explain (S*olutionhood*) F : RS	B	B	B		B
1,2,1	15	R-ana, P, C	B	U: OR, explain (*Justify*) F : merge, replace, explain (*Justify*)	B	B	B	B	B
2,3,2	20	R-syn, A,C	M	U : OR, explain (*Justify, Introduction*) F : delete, merge, replace, explain (*Justify*)	M	M	B		B
3,4,3	30	R-ano, A, L	B	U : RS F : explain (*Justify, Evaluation, Purpose*)	B	B	B		B
4,2,1	25	R-syn, A,C	M	U : OR, explain (*Introduction, summary, Justify*) F : merge, replace, explain (*Justify*)	M	B	B		M
5,5,8	90	R-ana, A, L	B	NU : add,delete,reorder,explain (*Solutionhood*) F : replace, explain (*Justify, Background, Evaluation*)	B	M	M		B
6,2,1	35	R-syn, A,L	B	U : merge, replace, explain (*Justify, Summary, Introduction*) F : replace, explain (*Justify)*	B	B	B		B
7,4,3	20	R-ano, A, C	B	U : RS, merge F : explain (Justify, *Evaluation*), merge	B	M	B		M
8,5,8	35	R-syn, A,C	B	NU : add, explain (*Solutionhood*) F : merge, delete, replace, explain (*Justify)*	B	B	M		B
8,1,1	50	R-ano, A, L	B	NU : delete,reorder,add,explain (*Solutionhood)* F : explain (*Justify, Background, Elaboration*)	B	B	M	B	B
9,6,10	20	R-syn, A,C	B	U : merge, replace, explain (I*ntroduction, Summary*) F : merge, replace	B	B	B		B
9,3,2	15	R-ana, A, C	B	U : Replace, delete, merge F : replace, delete	B	B	B	B	B
10,6,10	40	R-ano, A, C	B	NU : add, reorder F : merge, replace, explain (*Justify, Purpose, Background*)	B	B	M		B

selon un style personnel. Dans un rapport d'analyse, le participant se concentre plutôt sur le raisonnement de l'utilisateur tracé Dans un rapport d'analyse, le participant se concentre plutôt sur le raisonnement de l'utilisateur tracé et son comportement au moment où il a effectué l'activité tracée ; il reformule et enrichi ce contenu selon son propre style.

– La quatrième colonne présente une évaluation de l'utilité de l'outil (B : bon, M : moyen, F : faible) à travers l'évaluation des participants du document produit automatiquement (D_0) et de l'aptitude de l'outil à répondre à leurs besoins, ainsi que des informations sur l'utilisation des opérations supportées en distinguant celles utilisées (U), celles fréquemment utilisées (F) et celles non utilisées (NU) ; nous distinguons également les opérations rhétoriques (RS : comme *add* et *explain* segment) de celles organisationnelles (OR : comme *replace* et *merge* segments).

– La cinquième colonne présente une évaluation de l'utilisabilité de l'outil (B, M ou F) à travers la mesure de la facilité de sa première utilisation (apprenabilité), de la facilité de reprendre son utilisation après une période de non utilisation (Mémorabilité), de l'efficacité de son utilisation et de la satisfaction des participants à propos de cette utilisation.

7.3. Discussion

Evaluation de l'outil. Sept participants avaient une facilité à utiliser l'outil et une rapidité à accomplir leurs tâches ; alors que huit participants ont considéré que l'outil répondait vraiment à leurs besoins. Tous les participants ont apprécié le document (D_0) produit automatiquement et l'ont trouvé plus compréhensible que la trace d'activité dont il est issu (fichier log). Par ailleurs, deux participants (2 et 4) ont demandé de réduire et d'améliorer le contenu du document initial. De façon similaire, deux participants (2 et 7) ont proposé d'améliorer l'interface de l'outil en remplaçant le panneau central affichant les segments du document par un graphe de segments et de relations. Des participants (4 et 7) ont également réclamé de nouveaux besoins, en particulier la possibilité d'enrichir le contenu du document texte produit avec des données d'autres formes (vidéos, illustrations, etc.). Finalement, trois participants (5, 2, 10) avaient des difficultés à appliquer quelques opérations car ils n'avaient pas bien compris les principes des relations rhétoriques. Nous prévoyons de résoudre ces problèmes dans la nouvelle version de notre outil. Nous précisons que l'outil ActRedoc peut servir à redocumenter toute activité informatique à condition que celle-ci soit tracée et que ses traces puissent être modélisées selon notre modèle générique de trace (figure 3).

Durée du processus de redocumentation. Cette durée dépendait de la longueur de la trace redocumentée (cas du participant-8 ayant redocumenté la trace la plus longue et celle la plus courte), mais il ne s'agissait pas du seul facteur (cas du participant-5 qui a pris beaucoup de temps pour redocumenter la trace la plus courte). D'autres facteurs, tel que l'objectif du participant (cas des participants 8 et 5 qui ont redocumenté la trace-5) et/ou les performances de celui-ci (cas des participants 4 et 6 qui ont redocumenté la trace-2 selon des objectifs similaires) ont

influencé cette durée. Nous avons constaté que l'utilisateur tracé prenait moins de temps à redocumenter sa trace d'activité que d'autres participants (cas du participant-1 avec la trace-2) ; toutefois, cela ne peut être généralisé (cas des participant 9 et 2 avec la trace-3) du fait que le degré d'engagement dans le processus peut varier d'un participant à l'autre. Nous avons constaté également que le temps de la redocumentation de la trace d'une activité récente était plus court que celui de la trace d'une activité ancienne (cas du participant-1), ce qui peut se justifier par l'effet de la mémoire du participant sur la durée du processus.

Utilisation des opérations. Les utilisateurs tracés nous ont paru plus engagés dans le processus de redocumentation de leurs traces que les autres participants. Leur consultation des documents produits initialement (D_0) a montré qu'ils considéraient ces documents comme des supports pertinents à leurs mémoires. Durant le processus, ces participants ont utilisé différentes opérations (rhétoriques et organisationnelles) qu'ils avaient choisies avec soin (cas des participants 1, 2 et 8) malgré la longueur de la trace. En général, les participants étaient plus familiarisés aux opérations organisationnelles qu'à celles rhétoriques. Pour cela, certains d'entre eux (participants 3,5 et 7) ont consulté l'aide sur ces opérations à plusieurs reprises, alors que d'autres (participants 2,4 et 9) n'en ont utilisé que quelques-unes (particulièrement *explain* avec *Justify*, *Introduction* et *Summary*) en combinaison avec l'opération *replace.* L'un des participant (4) a proposé que l'ensemble des opérations rhétoriques soit réduit, mais un autre (1) a considéré que toutes les opérations supportées étaient utiles pour la redocumentation de son activité. Les opérations organisationnelles *replace* et *merge* étaient utilisées par la majorité des participants ; contrairement à *delete* et *reorder*, en particulier pour les participants qui ont redocumenté les traces des autres et qui ont hésité à falsifier leurs contenus (cas des participants 7 et 8). De plus, l'objectif de la redocumentation a influencé l'utilisation des opérations supportées par notre outil. Pour produire des rapports de synthèse, les opérations les plus utilisées étaient *merge*, *replace* et *explain (*avec *justify*) ; pour des rapports d'analyse, les opérations les plus utilisées étaient e*xplain* (avec *purpose*, *background* ou *elaboration)* ; alors que pour des rapports annotés, différentes opérations rhétoriques étaient utilisées.

8. Conclusion et perspectives

La documentation de l'activité médiée informatiquement est un processus qui nécessite la collecte de données décrivant cette activité. Une bonne partie de ces données résulte de l'enregistrement automatique de l'interaction de l'utilisateur avec le système informatique (traces d'activité), alors que l'autre partie est souvent fournie via une intervention humaine pour décrire le contexte de l'activité et expliciter la logique et les intentions de l'utilisateur tracé. En outre, collecter des données à propos de l'activité tracée n'est pas suffisant pour décrire celle-ci ; un processus interprétatif et organisationnel est notamment requis pour exprimer juste ce qui est important dans une forme compréhensible, cohérente et facilement échangeable. Dans cet article, nous avons présenté une approche semi-automatique de redocumentation en texte de l'activité informatique, basée sur la théorie de la

structure rhétorique (RST). Cette approche générique utilise des modèles formels pour les données d'entrée (trace d'activité) et de sortie (document produits) du processus de redocumentation et supporte celui-ci à travers la construction d'une narration dans une forme textuelle familière. Cette construction est caractérisée par deux phases. La première phase est automatique et génère un document texte fragmenté à partir de la trace d'activité ; tandis que la deuxième est interactive et permet à l'utilisateur de s'approprier le contenu et la forme du document initialement produit selon ses besoins et ses choix, tout en maintenant la cohérence du document. Le but étant d'aboutir à un document final personnalisé et facilement échangeable.

Nous avons développé un outil de redocumentation, appelé ActRedoc, comme une première implémentation de notre approche. Cet outil générique exploite des traces d'activité modélisées sémantiquement (par une ontologie) et un outil de génération du langage naturel pour pouvoir supporter la génération automatique d'un document texte initial. La modélisation sémantique de ce document et la spécification d'opérations pouvant être appliquées à celui-ci à base de règles rhétoriques a permis de préserver la cohérence de celui-ci durant sa manipulation par l'utilisateur, ainsi que son export dans une forme exploitable aussi bien par des humains que par des machines (texte enrichi de métadonnées). Une première évaluation d'ActRedoc auprès d'un groupe d'utilisateurs a montré leur satisfaction à propos de l'utilisabilité et l'utilité de l'outil. Cependant, des besoins ont été signalés, particulièrement pour faciliter l'utilisation des opérations rhétoriques et l'enrichissement du document texte produit avec d'autres médias (photos, vidéos...).

Nous avons exploité différentes technologies du web sémantique : des ontologies OWL pour modéliser les traces et les documents, un outil de génération du langage naturel (NaturalOWL) à partir de l'ontologie de la trace, et RDFa pour enrichir le contenu du document final avec des métadonnées sur la structure rhétorique du document et sur les éléments de la trace d'activité qu'il décrit. Notre choix de RST est dû à sa simplicité et à la diversité de ses applications pour modéliser les contenus de documents et contrôler leur cohérence. Notre but était de maintenir, tout au long du processus de redocumentation, suffisamment d'informations exploitables automatiquement afin de pouvoir supporter ce processus et assister l'utilisateur à construire de façon naturelle une histoire de l'activité tracée. Notre travail futur se focalise sur l'amélioration de notre outil générique de redocumentation en texte, ainsi que sur son test pour d'autres activités informatiques, éventuellement collaboratives. Nous veillerons à faire ces tests auprès d'utilisateurs potentiellement engagés dans le processus afin d'évaluer réellement l'utilité de notre outil. Nous voudrions également offrir plus de liberté à l'utilisateur pour personnaliser la présentation du document final et enrichir sa forme avec d'autres médias (hypermédia).

Remerciements

Les auteurs tiennent à remercier le développeur de l'application « Presse Illustrée » Pierre-Yves Londron pour sa coopération ainsi que les relecteurs anonymes de cet article pour leur commentaires très détaillés.

Bibliographie

Admed M., Hoang H., Karim M. S., (2004). SemanticLIFE: A framework for managing information of a human lifetime. *Proceedings of the 6th International Conference on Information Integration and Web-based Applications and Services,* p. 1-10.

Anjewierden A., Efimova L., (2006). Understanding weblog communities through digital traces: a framework, a tool and an example. *Proceedings of OTM Confederated International Workshops*, Montpellier, France, p. 279-289.

Appan P., Sundaram H., Birchfield D., (2004). Communicating Everyday Experiences. *Proceedings of the 1st ACM workshop on Story representation, mechanism and context*, p. 17-24.

Bal M., (1997). *Introduction to the theory of Narrative*. University of Toronto Press, Toronto, Canada, 2nd edition.

Barrutieta G., Abaitua J., Diaz J., (2002). An XML/RST-based approach to multilingual document generation for the web. *Proceedings of the Natural Language, n° 29*, Italy, p. 247-253.

Bocconi S., Nack F., Hardman L., (2005). Using rhetorical annotations for generating video documentaries. *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME'05)*, p. 1070-1073.

Brooks K., (1997). Do Story Agents Use Rocking Chairs? The Theory and Implementation of One Model for Computational Narrative. *Proceedings of Multimedia 1996*, p. 317-328.

Bush V., (1945). As We May Think. *The Atlantic Monthly*, vol. 176, n°1, 1945, p. 101-108.

Byrne D., Jones G. J. F., (2009). Creating Stories for Reflection from Multimodal Lifelog Content. *Proceedings of Designing for Reflection on Experience*, USA, p. 09-14.

Byrne D., Jones G.J.F., (2009b). Digital life stories: narrating personal experience from SenseCam collections. *http://doras.dcu.ie/15122/1/SenseCam_WS__Poster_DByrne_ 1.7.pdf*

Champin P. A., Prié Y., Mille A., (2003). MUSETTE: Modelling USEs and Tasks for Tracing Experience. *Proceedings of WS5*, Norway, p. 279-294.

Christophides V., (1998). Electronic Document Management Systems. *http://www.ics.forth.gr/~christop/*

Choquet C., Iksal S., (2007). Modeling Tracks for the Model Driven Reengineering of a TEL System. *Journal of Interactive Learning Research*, vol.18, n°2, p. 161-184.

Djemal K., Soulé-Dupuy C., Vallés-Parlangeau C., (2009). Multistructured documents: from modelling to multidimensional analyses. *SciWatch Journal*, vol. 2, n°4, p. 12-24.

Dragunov A.N., Diettrich T.G., Johnsruhe K., McLaughlin M., Li L., Herlocker J.L, (2005). TaskTracer: A Desktop Environment to Support Multi-tasking Knowledge Workers. *Proceedings of the 10th Conference on Intelligent User Interfaces,* USA, p. 75-82.

Fisher W. R., (1989). *Human Communication as Narration: Toward a Philosophy of Reason, Value, and Action.* Columbia: University of South Carolina Press.

Frenzel K., Müller M., Sottong H., (2004). *Storytelling.* Das Harun-al-Raschid-Prinzip.

Galanis D., Androutsopoulos I., (2007). Generating Multilingual Descriptions from Linguistically Annotated OWL Ontologies : the NaturalOWL System. *Proceedings of the 11th European workshop on natural language generation*, Germany, p. 1-4.

Garlatti S., Iksal S., (2003). Declarative Specifications for Adaptive Hypermedia Based on a Semantic Web Approach. *Proceedings of the 9th Conference of UM,* p.81-85.

Gemmell J., Aris A., Lueder R., (2005). Telling Stories With MyLifeBits. *Proceedingsof IEEE International Conference on Multimedia & Expo*, p. 6-9.

Gemmell J., Bell G., Lueder R., (2006). MyLifeBits: a personal database for everything. *Communications of the ACM (CACM)*, vol. 49, n°1, p. 88-95.

Gray P., McLeod I., Draper S., Crease M., Thomas R., (2004). A Distributed Usage Monitoring System. *Proceedings of Computer-Aided Design of User Interfaces*, Portugal, p. 121-132.

Harper R., Randall D., Smyth N., Evans C., Heledd L., Moore R., (2007). Thanks for the Memory. *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it (Volume 2)*, Swinton, UK, p. 39-42.

IPTO, (2003). *Lifelog proposer information pamphlet*. Defense Advanced Research Projects Agency Information Processing Technology Office document PIP 03-30.

Kim S., Alani H., Hall W., Lewis P. H., Millard D. E., Shadbolt N, R., Weal M.J., (2002). Artequakt: Generating Tailored Biographies from Automatically Annotated Fragments from the Web. *Proceedings of the 15th European Conference on AI*, France, p.10-16.

Kelliher A., Davenport G., (2007). Everyday Storytelling: supporting the mediated expression of personal testimony. *Proceedings of HCI (4)*, p.926-933.

Laflaquière J., Settouti L. S., Prié Y., Mille A., (2006). A trace-based System Framework for Experience Management and Engineering. *Proceedings of the second International Workshop on Experience Management and Engineering*, UK, p. 10-18.

Lindley S.E., Randall D., Sharrock W., Glancy M., Smyth N., Harper R., (2009). Narrative, memory and practice. *Proceedings of the 23rd BCS conference on HCI*, p. 1-9.

Mann W., Thompson S., (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text journal*, vol. 8, n°3, p. 243-281.

Martin P., Alpay L., (1996). Conceptual Structures and Structured Documents. *Proceedings of the 4th International Conference on Conceptual Structures*, Australia, p. 145-159.

Nakasone A., Ishizuka M., (2006). SRST: A Storytelling Model Using Rhetorical Relations. *Proceedings of the 3rd International Conference on Technologies for Interactive Digital Storytelling and Entertainment*, Germany, p. 127-138.

Nanard M., Nanard J., (1989). MacWeb : un outil pour élaborer des documents. *Proceeding of WOODMAN'89, Workshop on Object Oriented Document Manipulation*, Rennes.

Naja-Jazzar H., de Silva N., Skaf-Molli H., Rahhal C., Molli P., (2009). OntoReST: A RST-based Ontology for Enhancing Documents Content Quality in Collaborative Writing. *Journal of Computer Science (INFOCOMP)*, vol.8, n°3, p.03-13.

Nelson K., (1989). *Narratives from the Crib*. Harvard University Press, Cambridge:Mass.

Nielsen J., Loranger H., (2006). *Prioritizing Web Usability*. New Riders Press.

O'Hara K., Tuffield M., Shadbolt N., (2009). Lifelogging: Privacy and empowerment with memories for life. *Identity in the Information Society journal*, vol.1, n°2, p.155-172.

Oulasvirta A., Blom J., (2008). Motivations in personalization behavior. *Interacting with Computers journal*, vol.20, n°1, p. 1–16.

Pédauque, R., (2006), *Le document à la lumière du numérique*, C&F Editions.

Ranwez S., Crampes M., (1999). Conceptual Documents and Hypertext Documents are two Different Forms of Virtual Documents. *www8 workshop on Virtual Documents*.

Rizzo P., Shaw E., Johnson W. N., (2002). An Agent That Helps Children to Author Rhetorically-Structured Digital Puppet Presentations. *Proceedings of the 6th International Conference on Intelligent Tutoring Systems*, San Sebastian, p. 903-912.

Settouti L. S., Prié Y., Marty J. C., Mille A., (2009). A Trace-Based System for Technology-Enhanced Learning Systems Personalisation. *Proceedings of the 9th IEEE International Conference on Advanced Learning Technologies*, Riga, Latvia.

Silva N. D., Henderson P., (2005). Narrative support for technical documents: Formalising rhetorical structure theory. *Proceedings of the international conference on Enterprise Information Systems*, p. 105-110.

Strahovnik S., Mecava B., (2009). Storytelling and Web 2.0 Services: A synthesis of old and new ways of learning. *The eLearning Papers journal*, n°15, p. 1887-1542.

Taboada M., Mann W. C., (2006). Applications of Rhetorical Structure Theory. *Journal of Discourse Studies*, vol.8, n°4, p. 567-588.

Taboada M., Mann W .C., (2006b). Introduction to Rhetorical Structure Theory. *http://www.sfu.ca/rst/05bibliographies/report.html.*

Thomas P. M., (2005). Unified Activity Management: Explicitly Representing Activity in Work-Support System. *Proceedings of the European Conference on Computer Supported Cooperative Work*, France.

Tuffield M. M., Harris S., Dupplaw D. P., (2006). Image annotation with Photocopain. *Proceedings of the 1st International Workshop on SW Annotations for Multimedia*, p 1-14.

Tuffield M. M., Millard D. E., Shadbolt N. R., (2007). Ontological Approaches to Modelling Narrative. *Proceedings of the 2nd AKT DTA Symposium*, UK.

Yahiaoui L., Prié Y., Boufaïda Z., (2008). Redocumentation des traces d'activité médiée informatiquement dans le cadre des transactions communicationnelles. *Actes des 19ème Journées francophones d'Ingénierie des Connaissances*, France, p. 197-209.

Reçu le 01/06/2011
Accepté le 19/04/2012

Biographie

Leila Yahiaoui *est doctorante en informatique à l'Université Mentouri de Constantine (Algérie), membre de l'équipe de recherche Système d'Information et Bases de Connaissances (SIBC) du Laboratoire d'Informatique Repartie (LIRE) de l'université Mentouri-Constantine et doctorante à l'université Lyon1 (France), membre de l'équipe Supporting Interaction and Learning by Experience (SILEX) du Laboratoire d'InfoRmatique en Images et Systèmes d'information (LIRIS). Ses travaux de recherche portent sur l'annotation sémantique de documents, la modélisation des traces des activités informatiques et leur redocumentation.*

Yannick Prié *est maître de conférences HDR en informatique à l'Université Claude Bernard Lyon1 (France) et membre de l'équipe de recherche Supporting Interaction and Learning by Experience (SILEX) du Laboratoire d'InfoRmatique en Images et Systèmes d'information (LIRIS). Sa recherche concerne les systèmes d'interprétation et la réflexivité, et se décline en deux activités principales, l'une sur les systèmes à base de traces modélisées, l'autre sur les systèmes de lecture active audiovisuelle et d'annotation au sein du projet Advene.*

Zizette Boufaida *est Professeur des Universités au département informatique à l'Université Mentouri de Constantine (Algérie). Elle est membre de l'équipe de recherche Système d'Information et Bases de Connaissances (SIBC) du Laboratoire d'Informatique Repartie (LIRE) de l'université Mentouri-Constantine et responsable du doctorat "Ingénierie des Systèmes d'Information". Ses travaux de recherche portent sur la problématique du web sémantique, le développement des ontologies et leur représentation en logique de description, ainsi que l'extraction des connaissances.*