

# Personnalisation par un système multi-agent de la navigation au sein d'un corpus documentaire

Zina El Guedria<sup>a</sup>  
zina.el\_guedria@insa-rouen.fr

Laurent Vercouter<sup>a</sup>  
laurent.vercouter@insa-rouen.fr

<sup>a</sup>Laboratoire d'Informatique, du Traitement de l'Information et des Systèmes,  
Normandie Université, Institut National des Sciences Appliquées de Rouen, France

## Résumé

*La recherche documentaire dans un corpus numérique s'apparente à un processus de navigation guidé par un besoin d'information d'un utilisateur. Cette navigation nécessite l'usage d'outils classiques de recherche d'information pour sélectionner des documents pertinents en fonction d'une requête, mais ils doivent être complétés par des mécanismes de personnalisation et d'adaptation capable de faire évoluer la représentation du besoin en fonction des spécificités d'un utilisateur, de sa navigation en cours ou du corpus considéré. Nous proposons dans cet article un modèle multi-agent pour la navigation personnalisée permettant d'intégrer des facteurs hétérogènes de personnalisation. Ce travail s'inscrit dans le cadre d'un projet régional (PlaIR 2.0) ayant pour objectif le développement d'une plateforme d'accès personnalisé à un corpus numérique de documents juridiques.*

**Mots-clés :** *Coordination, Evolution, Adaptation, Systèmes mixtes*

## Abstract

*Documentary research in a digital corpus is like a crawling process guided by a user need of information. Such crawlings require the use of traditional information retrieval tools to select relevant documents based on a query, but they must be complemented by personalization and adaptation mechanisms able to change the representation of need according to the specificities of a user, his current crawling or considered corpus. We propose in this article a multi agent model for personalized crawling that take into account heterogeneous personalisation factors. This work is part of a regional project (PlaIR 2.0) whose aim is developing a personalized access platform to a digital corpus of legal documents.*

**Keywords:** *Coordination, Evolution, Adaptation, Mixed systems*

## 1 Introduction

L'accès aux documents d'un corpus numérique soulève des problèmes liés à la recherche d'information, à la visualisation des résultats d'une requête et à la navigation entre les documents. Certains de ces problèmes sont similaires à ceux rencontrés dans le domaine de la recherche d'information sur le Web (par exemple le calcul de la pertinence de documents par rapport à une requête) mais d'autres sont spécifiques au fait que l'on s'intéresse à un corpus fermé concernant des types de documents, requêtes et utilisateurs restreints que l'on peut représenter de manière plus fine que dans le cadre général du Web. Dans ce cadre il est réaliste de pouvoir identifier et chercher à reconnaître des pratiques ou usages spécifiques à l'accès à une base documentaire donnée et d'adapter en conséquence le comportement de la plateforme logicielle.

Nous nous intéressons dans cet article à la personnalisation de la navigation dans un corpus de documents. Les facteurs à prendre en compte dans cette personnalisation sont multiples : un profil propre à l'utilisateur, un cas d'usage reconnu pour la navigation, une proximité sémantique entre documents ou une recommandation construite sur l'historique des navigations. Nous proposons de traiter cette diversité d'influence par un système multi-agent. Le besoin informationnel et l'état courant d'une navigation sont représentés dans notre modèle par un environnement virtuel partagé que les agents perçoivent et sur lequel ils peuvent agir. Cet environnement, implémenté dans une couche d'artefacts, est ainsi un objet construit et adapté par l'activité collective des agents et de l'utilisateur, qui représentent la couche décisionnelle. Les contributions décrites dans cet article sont :

- une modélisation sous la forme d'agents et d'artefacts d'un système multi-agent pour réaliser une navigation personnalisée ;
- une représentation d'un profil de navigation incluant un processus de reformula-

- un processus de recommandation de documents à base de profils proches.

Cet article est organisé comme suit : La section 2 donne un aperçu de quelques travaux en RI personnalisée. Plus particulièrement, nous décrivons quelques travaux utilisant les systèmes multi-agents pour la recherche d'information ainsi que des travaux en recherche d'information personnalisée. La section 3 présente notre modélisation multi-agent de la plateforme de navigation personnalisée décrite selon deux couches : une couche Navigation et une couche Décision. La section 4 présente la problématique applicative. Nous y présentons le corpus documentaire utilisé, la réalisation de la couche Navigation par des artefacts ainsi qu'un scénario d'exécution. Enfin, la conclusion résume les travaux décrits dans cet article ainsi que nos perspectives de travaux futurs.

## **2 Recherche d'information et systèmes multi-agent**

### **2.1 Système multi-agent pour la recherche d'information**

Les premières approches multi agents apparues dans le domaine de recherche d'information se placent dans le cadre de la modélisation et résolution des systèmes de recherche d'information. Généralement, les tâches du processus sont affectées à différents agents qui peuvent les exécuter en séquence ou en concurrence. Dans ces modèles multi-agent les auteurs ont utilisé des agents utilisateurs, des agents ressources et des agents de fonctionnements et des agents d'ontologie etc. Parmi les premiers systèmes multi agents pour la recherche d'information nous pouvons citer à titre d'exemple : InfoSleuth [1], le système NetSA [2], la librairie digitale UMDL [3], le modèle RETSINA [8] et le système AgentSeek [9].

Dans le contexte de personnalisation de la recherche d'information interviennent les systèmes multi agents tel est le cas de SARI-POD (Système multi-Agent de Recherche Intelligente POSSibiliste de Documents Web) [4]. Ce modèle est un modèle multi-agent qui offre une collaboration entre les différents acteurs et la mise en oeuvre de toutes les fonctionnalités du système de recherche d'information. Il est à base de deux Réseaux Petits Mondes Hiérarchiques (RPMH) et d'un Réseau Possibiliste

(RP) : Le premier RPMH consiste à structurer les documents retrouvés en zones denses de pages Web thématiquement liées les unes aux autres. Le second RPMH consiste à ne pas chercher seulement le mot-clé dans les pages Web mais aussi les substantifs qui lui sont sémantiquement proches. Les Réseaux Possibilistes combinent les deux RPMH afin d'organiser les documents recherchés selon les préférences de l'utilisateur.

Dans le même cadre nous citons la plateforme SWAPP (Search based Web AdaPtive Platform), la mise en place des systèmes interactifs auto-adaptatifs par systèmes multi-agents auto-organiseurs appliqués à la personnalisation de l'accès à l'information [10]. Ce système utilise l'approche de AMAS (Adaptive Multi-Agent System) pour proposer une évaluation adaptative et personnalisée du feedback implicite de l'utilisateur en utilisant l'UIM (User Interest Manager) ainsi que la construction adaptative de son profil à base de l'UPM (User Profile Manager) en utilisant des documents textuels représentant ses intérêts.

### **2.2 Recherche d'information personnalisée**

Le besoin de la personnalisation de la recherche d'information commence par la difficulté de l'expression de besoin. En effet, pour affiner leurs besoins d'information les utilisateurs n'arrivent pas à trouver les termes exacts exprimant leurs besoins [11] et même en faisant des efforts, le besoin reste toujours ambigu [12]. Cette tâche est difficile même pour les utilisateurs expérimentés [13] et se termine par le choix de la pertinence, sous ses différents types [14, 15, 16] et pour un seul type de pertinence différents facteurs existent [14].

Avec l'apparition de la problématique de personnalisation dans la recherche d'information plusieurs travaux sont apparus. Nous pouvons distinguer la personnalisation individuelle [17, 18] où l'utilisateur est considéré de manière isolée et le modèle utilisateur est construit sur la base de ses préférences de contenu ou activités et mobilité via les applications. La personnalisation collaborative [19, 20] où l'utilisateur est considéré comme membre d'un groupe d'utilisateurs et le modèle utilisateur est construit sur la base de ses préférences de contenu ou activités via les applications ainsi que les modèles des autres utilisateurs similaires.

Il existe trois modes de personnalisation à savoir : La personnalisation explicite, implicite

et hybride. Commençons par la personnalisation explicite. Dans ce cas, la collecte des données descriptives de l'utilisateur est fournie par l'utilisateur. Par exemple les données démographiques (âge, sexe, langue maternelle [21, 22]), le choix des préférences (langue, genre de documents [23, 18]), le sujet d'intérêt [24] et les jugements qualitatifs sur l'information (j'aime, j'aime pas) [25, 17]. Les données citées ci-dessus peuvent être collectées à travers des formulaires (cases à cocher, saisie de mots clés), des interfaces élaborées (expression d'exemples, contre exemples, votes, notes, annotations) ou des questionnaires.

En ce qui concerne la personnalisation implicite, les données sont collectées à partir des données démographiques qui peuvent être déduites des interactions et des activités des utilisateurs [26] ou des requêtes [27]. Nous citons aussi les données sociales de l'utilisateur telles que les annotations, les posts, les blogs, les messages, les signaux, etc. [30, 28, 29].

L'activité de l'utilisateur et une source primordiale pour la personnalisation implicite. Cette activité peut être déterminée d'une part à partir des pages visitées, des requêtes passées [31, 17], des données de navigation [33, 32], des applications utilisées [33, 34], des favoris [34] et de l'historique de localisations [35] de l'utilisateur. D'autre part des interactions de l'utilisateur telles que les mouvements des yeux ; les données de clics [36] ; les actions sur les documents (ouverture, fermeture, impression, temps de lecture etc.) [37] ; les messages (e-mails) envoyés ou reçus [38] ; les annotations sociales, et les bookmarks [28, 29].

Enfin la personnalisation hybride utilise à la fois la personnalisation implicite et la personnalisation explicite. Cette personnalisation hybride peut commencer par la collecte explicite puis implicite (système WAIR[39]) ou la collecte implicite puis explicite tout en validant la collection implicite par des retours de l'utilisateur.

Un autre axe de personnalisation traite la durée de la session de recherche à savoir, la personnalisation basée sur le contexte courant (court terme) et la personnalisation basée le contexte passé (long-terme). Pour la personnalisation à court terme, a mémoire de la personnalisation est égale à la durée de la session. Dans ce cas, on rencontre un problème de délimitation de la session de recherche. Soit une délimitation basée sur le temps (par exemple 30 min dans [40]), une délimitation basée sur la similitude de re-

quêtes successives [41, 42] ou une délimitation basée sur la similitude des résultats de recherche [43]. La personnalisation basée sur le contexte passé (long-terme) où la mémoire de personnalisation est égale à la durée du compte utilisateur. Dans ces systèmes, on peut délimiter les sessions de recherche [18, 7] ou ne pas les délimiter [17].

### 3 Système multi-agent pour une navigation personnalisée

Cette section décrit le système multi-agent proposé pour réaliser une navigation personnalisée dans un corpus numérique de documents. Une approche multi-agent a été adoptée de manière à représenter l'hétérogénéité des facteurs de personnalisation à appliquer. Chaque agent exerce ainsi une influence différente sur la sélection de documents à présenter à l'utilisateur. La première partie de cette section présente l'architecture globale du système avant que soient détaillés l'environnement partagé des agents, représentant la navigation en cours, puis les spécifications des agents pour la couche décisionnelle.

#### 3.1 Architecture multi-agent

L'objectif de notre plateforme est de permettre à des utilisateurs de naviguer dans un corpus fermé en visualisant séquentiellement différents documents et en affinant ou adaptant leur requête au fur et à mesure de leur session de navigation. Pour cela, il est nécessaire dans un premier temps d'utiliser des mécanismes classiques de recherche d'information pour sélectionner des documents pertinents en fonction d'une requête. Nous proposons de compléter ces outils par des mécanismes de personnalisation et d'adaptation pour faire évoluer la représentation du besoin. Cette évolution est dynamique car réalisée pendant une navigation en fonction du profil des utilisateurs, de leurs actions et des navigations précédemment observées. Pour cette raison, nous utilisons le terme de *besoin informationnel* pour désigner l'objectif global que la plateforme doit satisfaire plutôt que celui de *requête*, propre à une recherche ponctuelle d'information.

Nous représentons le besoin informationnel au sein d'un environnement virtuel, partagé par les agents. Ce besoin, initialement exprimé par une requête composée d'un ensemble de termes, est modifié par le biais d'actions des agents sur cet

environnement. L'évolution du besoin informationnel est ainsi le résultat d'un processus de co-construction impliquant les agents et l'utilisateur dans le but d'intégrer différentes sources de personnalisation et un contrôle par l'utilisateur. De plus, l'environnement partagé des agents inclut des outils nécessaires pour la navigation (index, moteur de recherche d'information, interface avec l'utilisateur), ainsi que l'ensemble des documents jugés pertinents à l'instant courant de la navigation.

Afin de distinguer la couche décisionnelle et la couche opérationnelle de la plateforme, nous avons opté pour une approche Agent-Artefact [44]. Les décisions d'évolution du besoin informationnel sont prises par le système multi-agent, parfois en interaction avec l'utilisateur. Le stockage des informations liées à la navigation et l'exécution de requêtes ponctuelles sont à la charge d'artefacts de l'environnement des agents. Cette architecture est représentée par la figure 1.

Au niveau de la couche *Navigation*, cinq types d'artefacts sont utilisés. Un artefact *Interface* encapsule les fonctions d'interaction directe avec l'utilisateur. Chaque instance d'artefacts de type *Navigation* représente le besoin informationnel d'une session de navigation pour laquelle une instance d'artefact *Recherche* est créée pour exécuter une recherche d'information sur ce besoin. Le résultat d'une recherche est stocké dans un artefact de type *Document*. Enfin, les artefacts de type *Profil* collectent des informations sur le comportement des utilisateurs.

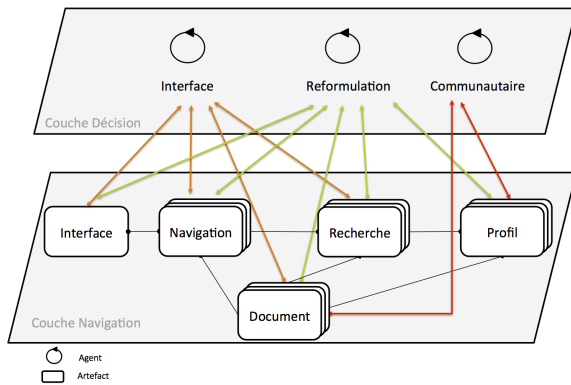


FIGURE 1 – Architecture Agent-Artefact proposée

La couche *Décision* contient les agents qui vont faire évoluer le besoin informationnel et

les documents présentés pendant la navigation. L'agent *Interface* sert de point d'entrée à l'utilisateur pour créer une requête initiale et la modifier pendant la navigation. L'agent *Reformulation* propose l'ajout de termes au besoin informationnel sur la base des documents sélectionnés par les recherches précédentes. Enfin l'agent *Communautaire* propose d'ajouter des documents jugés pertinents d'après l'historique des navigations passées similaires.

### 3.2 Couche navigation

Les cinq artefacts de la couche navigation sont spécifiés dans cette sous-section.

Les navigations ont lieu au sein d'un corpus numérique de documents. Nous notons  $\mathcal{D}$  l'ensemble des documents du corpus. Les requêtes sont effectuées par un ensemble de mots-clés parmi l'ensemble  $\mathcal{K}$ . Parmi ces mots-clés, un sous-ensemble  $\mathcal{T} \subset \mathcal{K}$  regroupe l'ensemble des termes utilisés pour indexer les documents du corpus. Enfin, nous faisons l'hypothèse que les termes indexant un document sont accessibles par une fonction *Index* telle que :

$$Index : \mathcal{D} \mapsto \mathcal{P}(\mathcal{T})$$

#### Artefact Navigation.

Un artefact Navigation encapsule un besoin informationnel d'un utilisateur. Soit  $\mathcal{N} = \{N_1, \dots, N_m\}$  l'ensemble de toutes les navigations, pour chaque navigation  $N_i$ , un utilisateur  $u_{N_i}$  a une requête courante  $K_{N_i}$  composée de plusieurs termes  $k_{N_i}^j$ .

$$N_i = (K_{N_i}, u_{N_i})$$

Avec

$$K_{N_i} = \{k_{N_i}^1, \dots, k_{N_i}^l\}$$

**Artefact Document.** Un artefact Document contient une référence aux documents jugés pertinents pour une navigation donnée  $N_i$ . Cet ensemble est amené à évoluer en cours de navigation en fonction des actions des agents et de l'utilisateur. L'ensemble de documents  $D_i$  correspondant à la navigation  $N_i$  est l'union des documents résultant d'une recherche d'information  $D_{i,RI}$  et de ceux recommandés par l'agent Communautaire  $D_{i,REC}$  :

$$D_i = (D_{i,RI}, D_{i,REC})$$

Avec :

$$D_{i,RI} = \{d_{i,RI}^1, \dots, d_{i,RI}^n\},$$

$$D_{i,REC} = \{d_{i,REC}^1, \dots, d_{i,REC}^p\}$$

**Artefact Profil.** Le profil  $P_i$  d'un utilisateur impliqué dans une session de navigation  $N_i$  est composé par sa requête initiale  $K_{P_i}^{INIT} \subset \mathcal{K}$ , les termes proposés en reformulation  $T_{P_i}^{ACC} \subset \mathcal{T}$  (acceptés par l'utilisateur) et  $T_{P_i}^{REJ} \subset \mathcal{T}$  (rejetés par l'utilisateur) et enfin les références des documents jugés pertinents par l'utilisateur  $D_{P_i}^{REL} \subset D_i$ .

$$P_i = (K_{P_i}^{INIT}, T_{P_i}^{ACC}, T_{P_i}^{REJ}, D_{P_i}^{REL})$$

avec

$$K_{P_i}^{INIT} = \{k_{P_i}^1, \dots, k_{P_i}^n\}$$

$$T_{P_i}^{ACC} = \{t_{P_i}^{ACC,1}, \dots, t_{P_i}^{ACC,p}\}$$

$$T_{P_i}^{REJ} = \{t_{P_i}^{REJ,1}, \dots, t_{P_i}^{REJ,q}\}$$

$$D_{P_i}^{REL} = \{d_{P_i}^{REL,1}, \dots, d_{P_i}^{REL,r}\}$$

et

$$K_{P_i}^{INIT} \cap T_{P_i}^{ACC} \cap T_{P_i}^{REJ} = \emptyset$$

### Artefact Recherche.

L'artefact Recherche sert de moteur de recherche d'information. Nous utilisons pour notre plateforme le moteur *Lucene*<sup>1</sup> qui est une application open source gratuite pour la recherche plein texte et l'analyse de contenu textuel. L'objectif de l'artefact est de prendre en entrée un ensemble de termes issus d'une requête et de fournir les références de documents jugés pertinents en sortie.

$$Recherche : (K) \mapsto P(D)$$

### Artefact Interface.

Le rôle de l'artefact Interface est de gérer les interactions avec l'utilisateur. Il récupère les requêtes ainsi que les retours de l'utilisateur pour le compte des agents Interface et Reformulation, et affiche les résultats des recherches effectuées par les agents.

1. <http://lucene.apache.org/core/>

## 3.3 Couche Décision

La couche décision est composée de trois agents qui ont la capacité de percevoir et d'agir sur leur environnement partagé représentant un processus de navigation en cours.

### Agent Interface.

Il existe un agent Interface par utilisateur, en charge du suivi de sa navigation. Un utilisateur  $u$  commence une navigation en saisissant une requête  $K_u = \{k_u^1, \dots, k_u^n\}$  composée d'un ensemble de mots-clés. Si l'utilisateur n'avait pas de navigation en cours, une nouvelle session de navigation est créée avec cette requête.

S'il existe déjà des navigations pour cet utilisateur, l'agent interface doit déterminer s'il s'agit du début d'une nouvelle navigation ou si l'insertion de nouveaux termes correspond à la continuité d'une session en cours. Dans ce but, l'agent calcule une mesure de similarité  $Sim(K_u, K_{N_i})$  entre les termes de la requête et ceux de toutes les navigations précédentes de l'utilisateur. Nous faisons ici l'hypothèse qu'une telle fonction existe sans nous restreindre à un calcul précis de similarité (pour avoir des exemples de similarité entre requêtes, voir [41, 42]). Si la similarité est en dessous d'un seuil  $\theta$  pour toutes les navigations passées, l'agent considère qu'une nouvelle navigation commence, sinon il considère être dans la continuité des plus similaires.

$$\exists N_i \in \mathcal{N}/u = u_{N_i}, Sim(K_u, K_{N_i}) > \theta,$$

$$K_{N_i} = K_{N_i} \cup K$$

Sinon, un nouvel artefact de navigation est créé :

$$\nexists N_i \in \mathcal{N}/u = u_{N_i}, Sim(K_u, K_{N_i}) < \theta,$$

$$N_m = (K, u)$$

À chaque fois qu'un artefact de navigation est créé ou mis à jour (notons-le  $N_j$ ) l'agent effectue une recherche par l'artefact correspondant  $Recherche(N_j)$  avec pour résultat de créer ou de mettre à jour l'artefact Document avec un ensemble  $D_{N_j,RI}$ .

Enfin, l'agent Interface envoie l'ensemble de documents  $D_j$  à l'artefact Interface pour les visualiser.

### Agent Reformulation.

Un agent Reformulation est associé à chaque utilisateur de la plateforme. Le fonctionnement de l'agent Reformulation est basé sur l'observation des artefacts. Il réagit à la perception qu'une navigation  $N_i$  et un ensemble de documents  $D_i$  ont été créés. Dans ce cas la première action de l'agent Reformulation est de créer un artefact Profil :

$$P_i = (K_{N_i}, \emptyset, \emptyset, \emptyset)$$

La deuxième partie du fonctionnement de l'agent Reformulation se base sur le retour de pertinence de l'utilisateur depuis l'artefact Interface. Dans notre système nous avons choisi le feedback explicite en demandant à l'utilisateur d'appuyer sur un bouton "c'est pertinent" pour exprimer son jugement de pertinence. L'ensemble des documents sélectionnés comme pertinents constitue l'ensemble  $D_{P_i}^{REL}$  qui est ajouté à  $P_i$ .

La proposition de reformulation se base sur les termes fréquents indexant les documents jugés pertinents qui sont obtenus par  $Index(D_{P_i}^{REL})$ . L'ensemble  $Proposition(N_i)$  contenant un ensemble de termes proposés par l'agent Reformulation à l'utilisateur est construit tel que :

$$\begin{aligned} \forall d \in D_{P_i}^{REL}, \forall t \in Index(d) / \\ t \notin K_{P_i}^{INIT} \cup T_{P_i}^{ACC} \cup T_{P_i}^{REJ}, \\ t \in Proposition(N_i) \end{aligned}$$

Cet ensemble est proposé à l'utilisateur qui peut choisir de les accepter (ensemble  $T_{acc}$ ) ou de les rejeter (ensemble  $T_{rej}$ ). Le profil est mis à jour par l'agent Reformulation avec

$$\begin{aligned} T_{P_i}^{ACC} &= T_{P_i}^{ACC} \cup T_{acc} \\ T_{P_i}^{REJ} &= T_{P_i}^{REJ} \cup T_{rej} \end{aligned}$$

Si  $T_{P_i}^{ACC}$  a été modifié, l'agent Reformulation lance une nouvelle recherche en utilisant la nouvelle requête :

$$Recherche (K_{P_i}^{INIT} \cup T_{P_i}^{ACC}) \mapsto D'$$

puis met à jour l'ensemble des documents pertinents à base de la recherche d'information.

$$D_i = (D_{i,RI} \cup D', D_{i,REC})$$

Ce processus de reformulation, ici décrit après le premier résultat d'une requête, sera réitéré à chaque fois que l'utilisateur sélectionnera des documents proposés comme étant pertinents (ceux-ci pouvant avoir été issus d'une requête déjà re-formulée).

### Agent Communautaire.

La couche décision contient un unique agent Communautaire dont le rôle est de proposer des documents qui ont été jugés pertinents pour des profils de navigation proche. À la création d'un profil  $P_i$ , l'agent Communautaire analyse tous les autres Profils  $P_j$  pour comparer leurs termes initiaux et acceptés à ceux de  $P_i$ . Si tous les termes de  $P_i$  sont inclus dans ceux d'un profil  $P_j$ , les documents jugés pertinents pour  $P_j$  seront recommandés pour  $P_i$ . La construction de l'ensemble  $D_{AJ}$  représentant l'ensemble des documents à ajouter à ceux recommandés se fait comme suit :

$$\begin{aligned} \forall P_j \in \mathcal{P} / K_{P_i}^{INIT} \subset K_{P_j}^{INIT} \cup T_{P_j}^{ACC}, \\ D_{P_j}^{REC} \subset D_{AJ} \end{aligned}$$

$$D_j = (D_{j,RI}, D_{j,REC} \cup D_{AJ})$$

## 4 Application à la navigation dans un corpus de document

La plateforme multi-agent de navigation personnalisée, modélisée dans la section précédente, a pour but d'être implémentée pour un prototype d'accès à un corpus numérique du droit international du transport. Nous décrivons dans cette section, le corpus utilisé pour le prototype, l'implémentation en cours de réalisation avec la plateforme Cartago [45], puis un scénario d'utilisation.

### 4.1 Le corpus de l'IDIT

Un prototype de plateforme de navigation sur un corpus numérique dans le domaine du droit international du transport est en développement dans le cadre du projet PlaIR 2.0 (Plateforme d'Indexation Régionale) en collaboration avec l'Institut du Droit International du Transport (IDIT).

Le corpus, constitué par l'IDIT [5], contient autour de 40 000 références incluant des documents classés en quatre groupes : jurisprudence,

articles, réglementation, fonds documentaires. Ces documents sont indexés à l'aide d'une terminologie du domaine et par une analyse plein texte des documents. La personnalisation en fonction des utilisateurs et des navigations est très intéressante dans ce type de corpus. D'une part les utilisateurs n'ont pas les mêmes besoins pour une même requête du fait de niveaux d'expertises différents. D'autre part, des navigations "typiques" ont lieu, sans être précisément formalisées, selon des usages fréquents comme des études comparatives de cas, de jurisprudence, des recherches précises, . . .

## 4.2 Les artefacts

La modélisation des artefacts du système pour leur implémentation dans Cartago [45] est représentée par la figure 2. Un artefact est défini par ses opérations et ses propriétés observables. À titre d'exemple, nous citons quelques opérations de l'artefact Profil qui permettent aux agents de mettre à jour les informations qu'il contient : *SetOwner*, *UpdateQuery*, *SetTacc* et *SetTrej* permettent respectivement de modifier l'utilisateur, les termes de son besoin informationnel courant, les termes acceptés et rejetés. Par ailleurs, les propriétés observables donnent des informations aux agents. Par exemple pour l'artefact Recherche *currentQuery* et *currentResult* contiennent respectivement les mots-clés d'une requête et les documents obtenus en résultat.

La notion de liens entre les artefacts (Links dans Cartago) permet le partage des opérations entre les artefacts. Dans l'artefact Profil, l'agent peut accéder et ajouter les documents pertinents ainsi que les besoins d'informations de l'utilisateur *AddRelevantD*, *AddQuery*. Par manque d'espace, nous ne pouvons pas détailler ici tous les artefacts du modèle.

La couche décisionnelle est implémentée en Jason, ce qui nous permet d'utiliser la plateforme intégrée *JaCa* (Jason+Cartago).

## 4.3 Scénario d'exécution

Dans cette partie, nous décrivons un scénario de navigation au sein du corpus de l'IDIT.

**Requête initiale.** Supposons qu'un utilisateur, Paul, initie une navigation en cherchant toutes les décisions juridiques ayant impliquées des camions. Sa requête initiale

est  $K_{N_1}^{INIT} = \{camions\}$ . L'agent Interface crée un artefact Navigation défini par  $N_1 = (\{camions\}, Paul)$ .

L'agent Interface lance, via l'artefact Recherche, une première requête : *Recherche(camions)* qui renvoie deux documents  $D_{N_1,RI} = \{Decision_1, Decision_{10}\}$ . Il s'ensuit la création d'un artefact Document  $D_1 = \{\{Decision_1, Decision_{10}\}, \emptyset\}$ .

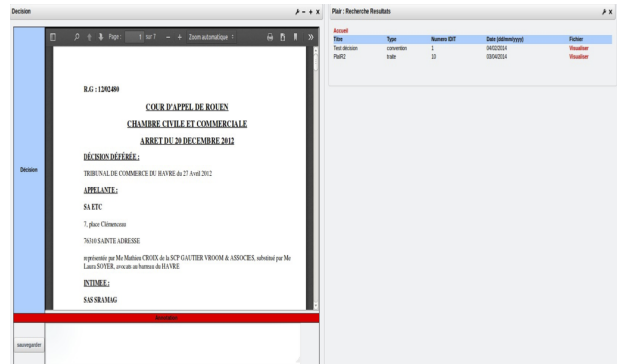


FIGURE 3 – Retour de pertinence

**Reformulations.** L'utilisateur choisit de visualiser les deux documents (Fig3). Mais, il juge la décision<sub>10</sub> comme un résultat pertinent. Ce retour de pertinence, ainsi que les premiers résultats, entraîne la création d'un artefact Profil par l'agent Reformulation initialisé à  $P_1 = (\{camions\}, \emptyset, \emptyset, \{Decision_{10}\})$ . L'agent Reformulation extrait les termes de l'index de la décision<sub>10</sub>. *Index(Decision<sub>10</sub>)* qui renvoie  $\{véhicules\}$  motorisés, camions, fruits, aliments, accidents, motocyclettes, automobiles.

Les termes ayant le poids le plus important et ne contenant pas les termes du besoin d'information initial sont : *véhicules motorisés*, *accidents* et *fruits*. L'agent Reformulation propose alors ces termes à l'utilisateur qui a le choix de les accepter ou les rejeter. Supposons, que Paul n'accepte que les deux derniers, son profil de navigation est alors devenu :  $P_1 = (\{camions\}, \{accidents, fruits\}, \{véhicules\} motorisés, \{Decision_{10}\})$ .

L'agent Reformulation lance une nouvelle recherche :

*Recherche(camions, accidents, fruits)* qui renvoie comme résultats les décisions (1, 10, 15, 20 et 19). L'utilisateur consulte les décisions (20, 19 et 15) et juge les décisions 15 et 19 comme pertinentes. L'agent Reformulation met alors à jour le profil de l'utilisateur *Paul*.



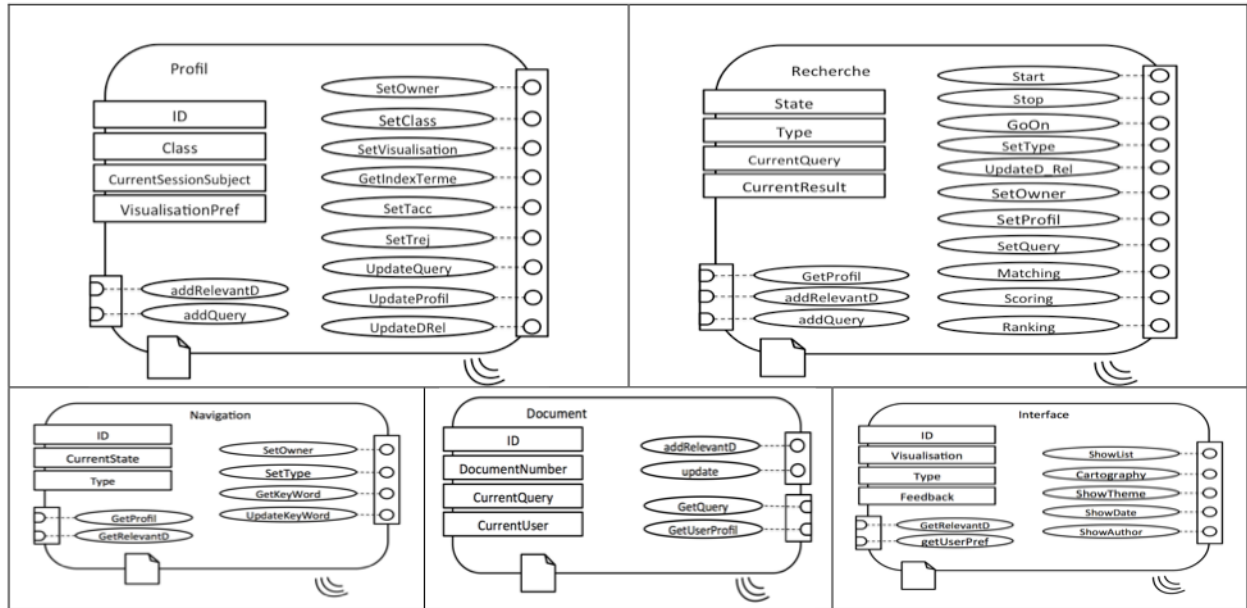


FIGURE 2 – Les artefacts

$P_1 = (\{\text{camions}\}, \{\text{accidents, fruits}\}, \{\text{véhicule motorisés}\}, \{Decision_{10}, Decision_{15}, Decision_{19}\})$ .

Les termes de l'index de la décision<sub>15</sub> et la décision<sub>19</sub> sont : *fruits, camions, dommages, marchandises, poids lourd, pourcentage de dommages* et *délai de livraison*. À partir de la proposition de ces termes et qui ne font pas partie du besoin d'information initial et des termes déjà acceptés ou rejetés, l'utilisateur sélectionne les termes qui lui semblent utiles pour affiner son besoin d'information, ce qui amène à la proposition de nouveaux documents que l'utilisateur pourra à nouveau juger comme pertinents. Ce processus continue jusqu'à ce que l'utilisateur n'accepte plus de termes pour affiner son besoin d'information ou s'il exprime un nouveau besoin qui caractérise une nouvelle navigation.

**Recommandation communautaire.** Supposons qu'un utilisateur, Jean, utilise la plateforme pour une navigation ultérieure à Paul. L'artefact de navigation associé est  $N_2 = (\{\text{fruits, délai de livraison}\}, \text{Jean})$ . Parallèlement aux documents proposés par les agents Interface et Reformulation, l'agent Communautaire compare ce besoin d'information initial avec les autres profils connus. Comme résultat il trouve  $\{\text{fruits, délai de livraison}\} \subset \{K^{INIT} \cup T^{ACC}, \text{Paul}\}$ . Il propose comme documents recommandés les documents jugés

pertinents par *Paul* pour ce besoin d'information à savoir les décisions (27, 15, 19 et 10). Le résultat à afficher à l'utilisateur est composé des résultats de l'agent Interface et des recommandations de l'agent Communautaire : Décision<sub>1</sub>, Décision<sub>51</sub>, Décision<sub>11</sub>, Décision<sub>19</sub> et Décision<sub>27</sub>, Décision<sub>15</sub>, Décision<sub>10</sub>, Décision<sub>19</sub>.

## 5 Conclusion et Perspectives

Nous avons présenté à travers cet article un modèle multi-agent pour accompagner la navigation dans un corpus numérique de documents. Le corpus de documents du Droit International du Transport, considéré dans le projet PlaIR 2.0, est représentatif d'un cas d'accès à l'information où la personnalisation est un aspect crucial à prendre en compte et qui comporte différents points de vue à savoir un profil propre à l'utilisateur, un cas d'usage ref pour la navigation, une proximité sémantique entre documents ou une recommandation construite sur l'historique des navigations.

Nous avons proposé une modélisation sous la forme d'une couche d'artefacts pour représenter par un environnement partagé la navigation des utilisateurs, et sous la forme d'une couche décisionnelle, les agents qui vont, conjointement avec l'utilisateur influencer la navigation en faisant évoluer leur environnement. Un scénario de navigation illustre de quelle manière les



agents vont pro-activement proposer des reformulations de requêtes ou des documents jugés pertinents lors de navigations passées. En perspectives, nous envisageons une validation expérimentale sur un panel d'utilisateurs.

## 6 Remerciements

Les travaux menés dans cet article bénéficient d'un financement du Grand Réseau de Recherche : Logistique, Mobilité, Numérique de la Région Haute-Normandie (projet PlaIR 2.0, 2013-2016).

## Références

- [1] Nodine, M., Fowler, J., Ksiezyk, T., Perry, B., Taylor, M., et Unruh, A. (2000). Active information gathering in infosleuthTM. *International Journal of Co-operative Information Systems*, 9(01n02), 3-27.
- [2] Marc Côté, Nader Troudi, *NetSA : Une Architecture Multiagent pour la Recherche sur Internet*, Expertise Informatique, Vol 3(3), 1998.
- [3] B. Chaib-draa, I. Jarras et B. Moulin, *Systèmes multi agents : Principes généraux et applications*, Edition Hermès (2001).
- [4] Elayeb Bilel, *SARIPOD : Système multi-Agent de Recherche Intelligente POSSIBILISTE de Documents Web*, Thèse de doctorat en informatique, Université de Toulouse, Toulouse (France), 2009.
- [5] *Institut du Droit International des Transports*, Site : <http://www.idit.asso.fr/>.
- [6] Gowan J., *A multiple model approach to personalized information access*, Master thesis in computer science, Faculty of science, Université de College Dublin, February, 2003.
- [7] Sieg A., Mobasher B., Lytinen S., Burke R., *Using Concept Hierarchies to Enhance User Queries in Web-based Information Retrieval*, Artificial Intelligence and Applications(AIA), 2004.
- [8] Katia Sycara and Keith Decker and Ananddeep Pannu and Mike Williamson and Dajun Zeng, *Distributed Intelligent Agents*, IEEE Expert, Vol. 11, pp. 36-46, 1996.
- [9] Grey D.J., Dunne G. and R.I. Ferguson, *A Mobile Agent Architecture for Searching the WWW*, In Proc. Workshop on Agents in Industry, 4th International Conference of Autonomous Agents, Barcelona, 2000.
- [10] Lemouzy, Sylvain, *Systèmes interactifs auto-adaptatifs par systèmes multi-agents auto-organisateurs : application à la personnalisation de l'accès à l'information*, Thèse de doctorat en informatique, Université Paul Sabatier - Toulouse III, Toulouse (France), 2011.
- [11] Jansen, B.J., Spink, A. and Saracevic, T. (2000). *Real life, real users, and real needs : A study and analysis of user queries on the web*. *Information Processing and Management*, 36(2), 207-227
- [12] Spink, A. et Jansen, B. (2004). *Web search : public searching of the Web*. Netherlands : Kluwer Academic Publishers.
- [13] Leake, D. B., et Scherle, R. (2001, January 14 - 17). *Towards context-based search engine selection*. Paper presented at the International Conference on Intelligent User Interfaces, Santa Fe, New Mexico, United States.
- [14] T. Saracevic. *Relevance : A review of the literature and a framework for thinking on the notion in information science*. part iii : Behavior and effects of relevance. *J. Am. Soc. Inf. Sci. Technol.*, 58(13) :2126-2144, 2007.
- [15] Borlund, P. (2003). *The concept of relevance in IR*. *Journal of the American Society for Information Science and Technology*, 54(10), 913-925.
- [16] Mizzaro, S. *Relevance : The whole history*. *JASIS*, 1997, vol. 48, no 9, p. 810-832.
- [17] Liu, F., Yu, C., et Meng, W. (2004). *Personalized web search for improving retrieval effectiveness*. *Knowledge and Data Engineering, IEEE transactions on*, 16(1), 28-40.
- [18] Sieg, A., Mobasher, B., et Burke, R. (2007, November). *Web search personalization with ontological user profiles*. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management* (pp. 525-534). ACM.
- [19] Teevan, J., Morris, M. R., et Bush, S. (2009, February). *Discovering and using groups to improve personalized search*. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining* (pp. 15-24). ACM.
- [20] Morris, M. R., Teevan, J., et Bush, S. (2008, November). *Enhancing collaborative web search with personalization : groupization, smart splitting, and group hit-highlighting*. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (pp. 481-484). ACM.
- [21] Hupfer, M. E., et Detlor, B. (2006). *Gender and Web information seeking : A self-concept orientation model*. *Journal of the American Society for Information Science and Technology*, 57(8), 1105-1115.
- [22] Weber, I., et Castillo, C. (2010, July). *The demographics of web search*. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 523-530). ACM.
- [23] Martin, I., et Jose, J. M. (2004, April). *Fetch : A Personalised Information Retrieval Tool*. In *RIAO* (Vol. 4, pp. 405-419).
- [24] Cheverst, K., Davies, N., Mitchell, K., Friday, A., et Efstratiou, C. (2000, April). *Developing a context-aware electronic tourist guide : some issues and experiences*. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 17-24). ACM.
- [25] Chen, L., et Sycara, K. (1998, May). *WebMate : a personal agent for browsing and searching*. In *Proceedings of the second international conference on Autonomous agents* (pp. 132-139). ACM.
- [26] Hu, J., Zeng, H. J., Li, H., Niu, C., et Chen, Z. (2007, May). *Demographic prediction based on*

- user's browsing behavior*. In Proceedings of the 16th international conference on World Wide Web (pp. 151-160). ACM.
- [27] Jones, R., Kumar, R., Pang, B., et Tomkins, A. (2007, November). *I know what you did last summer : query logs and user privacy*. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (pp. 909-914). ACM.
- [28] Karweg, B., Huetter, C., et Böhm, K. (2011, October). *Evolving social search based on bookmarks and status messages from social networks*. In Proceedings of the 20th ACM international conference on Information and knowledge management (pp. 1825-1834). ACM.
- [29] Hecht, B., Teevan, J., Morris, M. R., et Liebling, D. J. (2012). *SearchBuddies : Bringing Search Engines into the Conversation*. ICWSM, 12, 138-145.
- [30] Vallet, D., Cantador, I., et Jose, J. M. (2010). *Personalizing web search with folksonomy-based user and document profiles*. In Advances in Information Retrieval (pp. 420-431). Springer Berlin Heidelberg.
- [31] Shen, X., Tan, B., et Zhai, C. (2005, August). *Context-sensitive information retrieval using implicit feedback*. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 43-50). ACM.
- [32] Speretta, M., et Gauch, S. (2005, September). *Personalized search based on user search histories*. In Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on (pp. 622-628). IEEE.
- [33] Yau, S. S., Liu, H., Huang, D., et Yao, Y. (2003, November). *Situation-aware personalized information retrieval for mobile internet*. In Computer Software and Applications Conference, 2003. COMPSAC 2003. Proceedings. 27th Annual International (pp. 639-644). IEEE.
- [34] Dumais, S., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R., et Robbins, D. C. (2003, July). *Stuff I've seen : a system for personal information retrieval and re-use*. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval (pp. 72-79). ACM.
- [35] Ying, J. J. C., Lu, E. H. C., Lee, W. C., Weng, T. C., et Tseng, V. S. (2010, November). *Mining user similarity from semantic trajectories*. In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks (pp. 19-26). ACM.
- [36] Sun, J. T., Zeng, H. J., Liu, H., Lu, Y., et Chen, Z. (2005, May). *Cubesvd : a novel approach to personalized web search*. In Proceedings of the 14th international conference on World Wide Web (pp. 382-390). ACM.
- [37] Pretschner, A., et Gauch, S. (1999). *Ontology based personalized search*. In Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on (pp. 391-398). IEEE.
- [38] Bellotti, V., Begole, B., Chi, E. H., Ducheneaut, N., Fang, J., Isaacs, E., ... et Walendowski, A. (2008, April). *Activity-based serendipitous recommendations with the Magitti mobile leisure guide*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1157-1166). ACM.
- [39] Zhang, B. T., et Seo, Y. W. (2001). *Personalized web-document filtering using reinforcement learning*. Applied Artificial Intelligence, 15(7), 665-685.
- [40] Silverstein, C., Marais, H., Henzinger, M., et Moricz, M. (1999, September). *Analysis of a very large web search engine query log*. In ACM SIGIR Forum (Vol. 33, No. 1, pp. 6-12). ACM.
- [41] He, D., Göker, A., et Harper, D. J. (2002). *Combining evidence for automatic web session identification*. Information Processing et Management, 38(5), 727-742.
- [42] Jones, R., et Klinkner, K. L. (2008, October). *Beyond the session timeout : automatic hierarchical segmentation of search topics in query logs*. In Proceedings of the 17th ACM conference on Information and knowledge management (pp. 699-708). ACM.
- [43] Daoud, M., Lechani, L. T., et Boughanem, M. (2009). *Towards a graph-based user profile modeling for a session-based personalized search*. Knowledge and Information Systems, 21(3), 365-398.
- [44] Omicini, A., Ricci, A., Viroli, M. *Artifacts in the A&A meta-model for multi-agent systems*. Autonomous Agents and Multi-Agent Systems. Autonomous agents and multi-agent systems, 17(3), 432-456.(2008)
- [45] A. Ricci, M. Viroli, and A. Omicini. *CARtAgO : An infrastructure for engineering computational environments in MAS*. In D. Weyns, H. V. D. Parunak, and F. Michel, editors, Proc. of E4MAS, pages 102-119, Hakodate, Japan, 2006.