

# Définition de parcours sur un Web Communautaire

(e-parcours)

par [Philippe.Picouet@enst-bretagne.fr](mailto:Philippe.Picouet@enst-bretagne.fr) et [Jean-Marc.Saglio@enst.fr](mailto:Jean-Marc.Saglio@enst.fr)

---

## Sommaire:

1. [Introduction](#)
  2. [Web et portails communautaires](#)
  3. [Les itinéraires: un espace de découverte des ressources documentaires](#)
  4. [Modèles formels d'itinéraires](#)
  5. [Retour sur la composition des itinéraires: le parcours comme modèle alternatif](#)
  6. [Conclusion et perspectives](#)
  7. [Annexes et bibliographie](#)
- 

## 1. Introduction

Ce rapport présente succinctement les résultats du projet incitatif GET RE112, "e-parcours". Ce projet, indirectement connecté via un de ses partenaires (ENST Bretagne) à un projet européen ([Mesmuses](#), projet IST-2000-26274 [1]), avait pour ambition de mener une réflexion conjointe ENST et ENST Bretagne sur les problématiques actuelles du web sémantique (selon la [définition de Tim Berners Lee](#) [9]) et en particulier sur la réalisation de parcours dans le cadre plus restreint du web communautaire.

Le Web apparaît aujourd'hui partout dans le monde comme la plus large **base documentaire** en volume stocké et en accès distribué. Elle n'en est pas moins volatile, incomplète, hétérogène et même pleine de contradictions. L'analyse « full text » réalisée par les moteurs de recherche a vite montré ses limites et encouragé l'utilisation de métadonnées au sein des documents accessibles via le web. C'est ainsi que se sont succédées de nombreuses propositions pour encourager l'ajout de **métadonnées**, certaines juste syntaxiques (balise <META> de HTML), d'autres plus précises mais peu évolutives ([PICS](#) [19], [Dublin Core](#)[15]) et plus récemment des modèles extrêmement souples tels que [RDF](#) [13, 14]. En ajoutant des métadonnées aux documents, on souhaite rajouter au web la sémantique qui lui manquait.

Toutefois, l'interprétation des méta-données elle-même s'avère également d'une grande complexité. Aussi assiste-t-on depuis quelques années à l'apparition de portails qui présentent les sites web à travers des **schémas de connaissance** qui servent de **guide** pour des publics tantôt ciblés, tantôt pas. Des exemples de tels portails se trouvent dans les sites de l'[Open Directory Project](#) [20](navigation hiérarchique sur le modèle d'un **thesaurus**) et du [Semantopic Project](#) [21](navigation en réseau sur le modèle **topic-map**) . Les guides construits sur des modèles d'**ontologie** plus

complexes sont encore du domaine de la recherche, ce que l'on peut voir, par exemple, dans le site du [OntoBroker Project](#) [22].

Par ailleurs, depuis quelques années, se développe le concept de **web communautaire** alias **C-web** [23]. Il désigne un site dédié à un public spécialisé qui est supposé partager une **connaissance commune** dans sa spécialité. Exprimée et formalisée par un modèle d'**ontologie**, d'expression plus puissante que les simples thesauri ou topic-maps, cette connaissance devrait permettre une description plus riche par de multiples méta-données - une "indexation" - des documents d'intérêt reconnu pour la communauté. Ce qui permet d'espérer une manipulation - lecture et, peut être, (ré)écriture - du stock de documents ainsi communément décrits.

Parmi ces manipulations, la confection d'une suite ordonnée de ressources, instaurant une progressivité dans la lecture, introduit une approche résolument nouvelle et alternative aux outils classiques de recherche d'information (Information Retrieval) qui se focalisent davantage sur la recherche du bon document plutôt que sur la confection d'un nouveau document comme composition de documents existants.

Les questions que posent de telles fonctionnalités sont variées, tant du côté des prérequis techniques (mise en œuvre de modèles de représentation d'ontologies, d'architectures distribuées et d'outils de traitement adaptés) que des problématiques encore plus vastes concernant par exemple le degré de spécialisation de l'ontologie (ou la circonscription de la communauté concernée) qui optimiserait l'utilisabilité du web communautaire, tant du point de vue de l'indexation que de la recherche.

Les applications en sont très nombreuses: **e-commerce**, **e-formation**, etc... surtout dans les domaines des transports, du tourisme et tous les secteurs de l'information et de la culture (presse, universités, centres de recherche, bibliothèques, musées, archives ...).

Après une description synthétiques des principaux aspects techniques du web communautaire (section 2), nous nous efforcerons (section 3) de clarifier la problématique du parcours. Puis (section 4) nous formaliserons une conception générale des itinéraires. Enfin (section 5), considérant sur l'influence de l'indexation sur la réalisation des itinéraires, nous décrirons nos perspectives finales de recherche.

## 2. Web et portails communautaires

Fondé sur l'existence de communautés pouvant s'identifier par leur ontologie, le web communautaire mise sur la qualité des descriptions des ressources. Pour garantir cette qualité, il encourage une politique dite du « push », dans laquelle les utilisateurs sont encouragés à fournir eux-mêmes la description des documents. En se différenciant des politiques « pull » plus généralement utilisées par les moteurs de recherche, le web communautaire concentre deux défis :

- Le premier concerne la conception d'un outil susceptible de supporter la gestion de **repository** de description des objets du monde réel et des ressources du web, un outil à la fois suffisamment souple pour modéliser et interroger une ontologie rapidement complexe, et suffisamment robuste pour gérer le volume considérable de ces descriptions.

- Le deuxième concerne la mise en place d'une **architecture distribuée**, suffisamment souple pour ne pas renier l'apparement du web communautaire avec le web classique, suffisamment contrainte pour garantir la cohérence entre les ressources distantes sur le web et leur description dans un repository.

Dans cette section, nous abordons successivement les questions liées à la création des ontologies (modèles, outils, etc.) puis celles liées aux architectures propres au web communautaire.

## 2.1 Ontologies et modélisation des métadonnées

Modéliser des métadonnées n'induit pas nécessairement un modèle de données complexe. A titre d'exemple, le Dublin Core [15] se limite à un ensemble très restreint de 15 champs alphanumériques (TITLE, AUTHOR, etc.). Néanmoins, malgré son intérêt, le Dublin Core (avec son unique champ DESCRIPTION) reste extrêmement limité pour décrire le contenu d'un document.

Pour obtenir une description manipulable du contenu sémantique d'un texte, un tel modèle est évidemment insuffisant. En réponse à un tel besoin, le World Wide Web Consortium (W3C) a stabilisé dès 1999 la spécification RDF [13](Resource Description Framework), qui présente un modèle permettant de décrire des ressources, accessibles ou non via le web. Elle repose sur une modélisation par triplet (ressource, propriété, valeur) dans laquelle la valeur peut désigner une valeur textuelle (littéraires) ou un autre objet. Depuis, une recommandation a été proposée pour doter RDF de schémas (RDFS [14]) permettant de contraindre les descriptions RDF. Cette approche fait actuellement l'objet de beaucoup d'intérêt par la communauté bases de données qui voit dans cette proposition un consensus possible entre la souplesse exigée par l'indexation de documents et les performances exigées par les systèmes opérationnels destinés à traiter des gros volumes d'information et devant fournir des fonctionnalités avancées en terme de langages de requêtes, en particulier.

D'autres modèles existent, plus proches des recherches sur la représentation des connaissances, et souvent plus souples encore : Topic maps [21], OIL [12], etc...

## 2.2 Architecture de référence

Une architecture de référence pour un site c-web comportera toujours typiquement les éléments suivants:

- o un sous-système de stockage des documents intéressant la communauté faisant fonction d'**entrepôt**, réel ou virtuel, de ressources du web (accessibles par HTTP)
- o un sous-système de description de ces documents reliées à des descriptions des objets et propriétés formant le domaine commun de connaissance de la communauté; cette base de méta-données peut être implémentée avec un modèle structuré classique, par exemple SQL99, ou, ce qui est plus homogène avec le web sémantique vu par le W3C, avec un modèle semi-structuré comme RDF/XML; ce sous système fait fonction de **catalogue (repository** de descriptions), local ou distribué, du site communautaire

- un sous-système de génération automatique de guide pour le portail communautaire, faisant fonction d'outil de **publication d'itinéraires de visite** guidée dans le c-web.

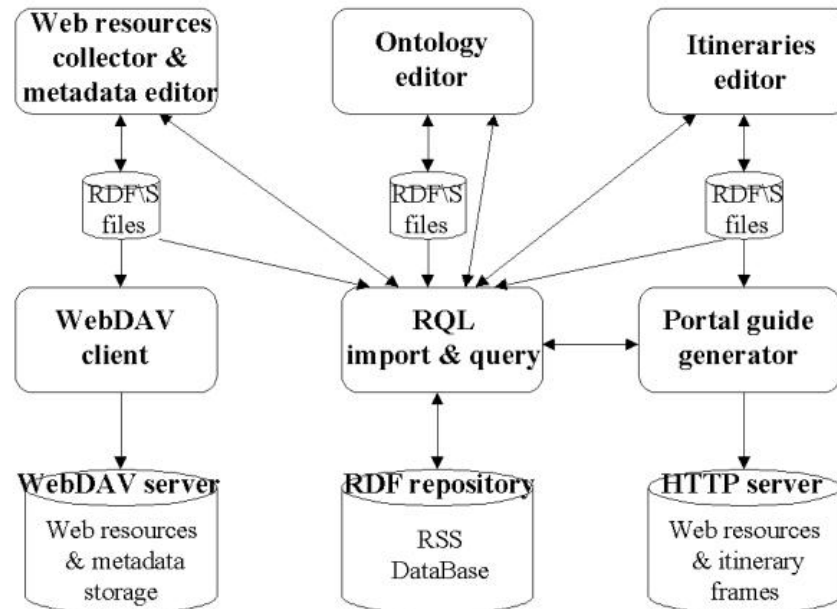


Figure 1: architecture de référence

Dans une telle architecture

- les documents n'entrent pas librement dans l'entrepôt. Ils font l'objet d'un processus d'acquisition selon le protocole client-serveur WebDAV destiné à assurer la cohérence entre la ressource du web et sa description (au moins celle qui correspond au "Dublin Core") et à en garder une copie, locale ou distribuée
- chaque document/ressource est ensuite associé - "**indexation**" - à un ou plusieurs objets ou collections d'objet de la base de connaissances communautaire définis par une requête (exprimée par exemple dans le RDF Query Language, [RQL](#) [18]) sur cette base (par exemple la RDF Schema Specific DataBase, [RSSDB](#) [17])
- la définition de cette base de connaissance peut évoluer ce qui peut entraîner une révision de l'indexation
- la définition d'un patron de publication (schémas de recherche, guides de visite, etc..) est toujours préalable à la génération automatique de portail.

### 3. Les itinéraires: un espace de découverte des ressources documentaires

**De l'IR à l'itinéraire** : e-parcours se distingue de la recherche d'information (IR) au sens classique du terme dans la mesure où son ambition est de proposer, non pas des documents candidats pour répondre (certes de manière la plus correcte possible) aux requêtes isolées d'un utilisateur, mais bien un itinéraire, sorte d'hypermedia (document composé/relié à d'autres documents), que l'utilisateur va parcourir et qui va lui donner « progressivement » les éléments de réponse à sa question principale. Cette progressivité résulte en partie d'une démarche argumentative.

Dans ce contexte, on peut prendre comme définition provisoire d'un itinéraire "*un chemin structuré sur des ressources documentaires*", cf. figure 2 ci-après.

Comme, à ce stade, tout reste à préciser sur la structure de cet itinéraire, on peut se demander si une **cartographie des documents** telle que celle suggérée par les représentations graphiques des ontologies - schémas de connaissance ou de métadonnées - paraît susceptible de faciliter, voire d'automatiser ce genre de réalisation.

Cette référence à la cartographie n'est pas innocente : si un rapprochement rapide par **la métaphore routière** est possible (elle est presque sous-entendue dans le terme d'itinéraire), il est intéressant d'en situer les limites : peut-on, comme la carte routière permet de le concevoir pour des itinéraires réels, utiliser la carte des concepts ontologiques pour concevoir des itinéraires entre documents ? Peut-on étendre certaines manipulations cartographiques classiques, telles le zoom, aux itinéraires électroniques ?

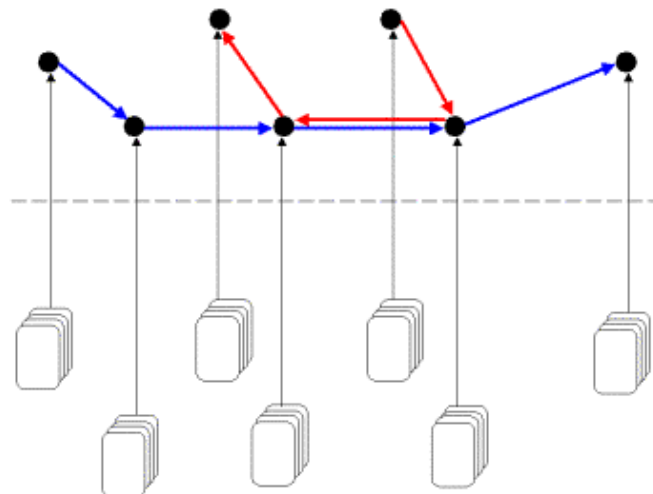


Figure 2: itinéraire sur des ressources documentaires

Si la carte routière et la carte sémantique sont toutes deux représentables sous la forme d'un graphe orienté, il apparaît rapidement que leurs sémantiques sont si différentes que leur utilisation à des fins de conception d'itinéraires sont difficile à concevoir :

- dans le graphe représentant la carte routière, les noeuds représentent des endroits spécifiés par leur localisation, leur position. Transféré dans un

contexte documentaire, la position d'un document renvoie à sa description au sein du graphe de métadonnées, mais ne sous-entend nullement, contrairement à la carte routière, une quelconque notion de distance. En outre, si, pour la carte routière, un lieu réel correspond de façon univoque à un lieu sur une carte, un document ne saurait quant à lui être mono-indexé. Inversement, un concept indexe évidemment plusieurs documents.

- dans le graphe représentant la carte routière, les routes représentent des transferts possibles d'un lieu à un autre et correspondent de manière univoque aux routes utilisées dans les itinéraires. En revanche, dans le contexte d'un itinéraire documentaire, les liens unissant les documents ne sont pas de nature ontologique mais plutôt de nature argumentative ou pédagogique : ils ne peuvent pas, a priori, se déduire des relations de l'ontologie.

Pour formaliser les itinéraires qui ne sont pas calqués sur le graphe des métadonnées, nous proposons de distinguer dans la représentation de notre modèle, cf. figure 3 ci-après, trois niveaux :

1. **knowledge support:** c'est le niveau documentaire à proprement parler, celui des ressources du web.
2. **knowledge representation:** c'est le niveau des métadonnées, qui sont des propriétés des objets d'un domaine de description, formant l'ontologie de ce domaine. Les documents sont indexés par ces métadonnées.
3. **knowledge transmission:** c'est le troisième et dernier niveau, qui décrit le graphe des itinéraires avec chaque étape reliée au graphe de métadonnées par une sorte de projection des étapes sur des objets et propriétés.

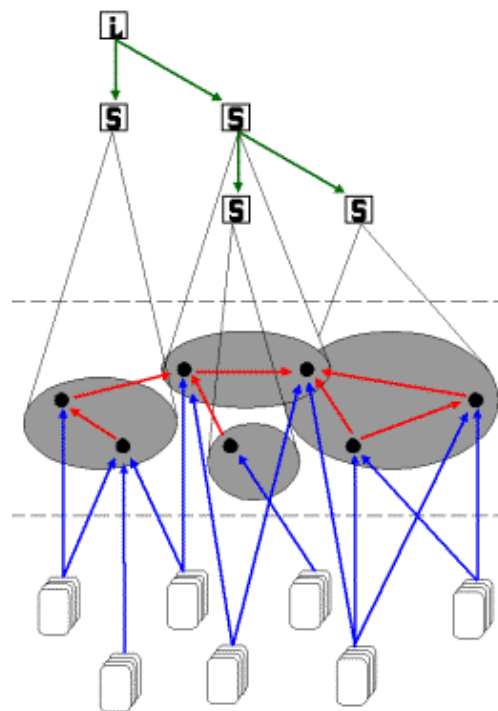


Figure 3: itinéraire sur des (collections d')objets référencés par des ressources documentaires

N.B. les descriptions reliées de bas en haut sur cette figure sont manipulés par les sous-sytèmes représentés de gauche à droite dans l'architecture de référence [plus haut](#).

## 4. Modèles formels d'itinéraires

### 4.1 Modèle hiérarchique avec prérequis

Dans la figure 2, ci-dessus, un itinéraire est vu comme une succession d'étapes. Ce modèle simpliste s'avère rapidement insuffisant en matière de structuration. Le passage à une structuration hiérarchique sur le mode étape – sous-étape (suivant la structuration classique des documents en partie – sous-partie), tout en suggérant un ordre par défaut (parcours de gauche à droite, en profondeur d'abord), offre une lisibilité plus grande. D'autres modèles sont néanmoins envisageables (voir section 4.3).

Nous proposons dans cette section un modèle formel d'itinéraire, qui découle directement de la structure de nombreux documents et rapports, à savoir une organisation hiérarchique que nous retrouvons dans la table des matières (TdM), et qui induit un ordre de lecture par défaut (parcours de l'arbre de gauche à droite, en profondeur d'abord). Cet ordre par défaut pouvant être librement contourné par l'utilisateur, les prérequis complètent cette modélisation en indiquant au lecteur (voire en l'obligeant) les sections à assimiler avant d'autres.

Un itinéraire est donc défini comme un triplet  $(S, P, I)$  dans lequel :

- $S$  est l'ensemble des étapes de l'itinéraire
- $I$  est un arbre ordonné dont les nœuds sont les étapes de  $S$  (par souci de simplicité, nous notons  $S_0$  le nœud racine de cet arbre).
- $P$  est un ensemble de prérequis représentés chacun par un couple  $(I_i, o_i)$  tel que :
  - $I_i$  est un sous ensemble de  $S$
  - $o_i$  est un élément de  $S$
  - $o_i$  n'est pas élément de  $I_i$

Le parcours de l'utilisateur au sein de ce graphe suit les règles suivantes :

- Chaque étape est triplement colorée : rouge si aucun de ses prérequis n'a été vérifié, orange si au moins un prérequis a été vérifié, et vert si le nœud a été visité.
- Initialisation : tous les nœuds sont rouges à l'exception des nœuds qui ne sont pas buts d'un prérequis. Ces nœuds-là sont orange (la racine  $S_0$  de  $I$  en fait partie).

A la première étape :

- localisation sur le nœud racine  $I$  (et visualisation du contenu associé)
- changement de la couleur de  $S_0$  à vert.
- calcul itératif des étapes permises :

- Pour chaque prérequis dont le but est rouge, si les étapes sources du prérequis sont vert ou oranges, alors changer la couleur de l'étape cible à orange.
- Recommencer tant qu'au moins une nouvelle étape devient orange à chaque itération

A chaque étape ultérieure :

- L'utilisateur peut se déplacer sur un nœud de son choix en fonction du mode de parcours qu'il a activé :
  - Parcours figé : il suit l'ordre linéaire imposé par l'arbre ordonné des étapes.
  - Parcours contraint : il peut aller dans n'importe laquelle des étapes oranges
  - Parcours libre : l'utilisateur peut aller dans n'importe quelle étape.
- Après chaque déplacement de l'utilisateur, on recalcule les étapes permises de la même manière qu'à la première étape.

D'autres pistes de recherche attirent actuellement notre attention. Nous les présentons ci-dessous en indiquant à chaque fois une utilisation pratique pour de tels travaux.

## 4.2. Modèle "rhétorique"

Une autre approche nous semble intéressante concernant la définition d'un modèle formel d'itinéraire. Elle est basée sur la théorie de la structure rhétorique [8]. Cette théorie, conçue pour l'analyse des liens argumentatifs au sein de documents, formalise leur représentation et dresse une liste de ces relations argumentatives et de leur effet sur le lecteur. Parmi ces relations, il semble qu'un petit nombre soit particulièrement bien adapté pour décrire par exemple la structure d'une exposition. Nous examinons actuellement l'intérêt d'utiliser ces relations argumentatives entre les étapes d'un itinéraires, l'objectif n'étant pas nécessairement de présenter ces relations aux utilisateurs mais de les utiliser pour générer plusieurs itinéraires « hiérarchiques ». Une étude allant dans ce sens est d'ailleurs en cours.

## 4.3 Modèle de méta-itinéraires

Les structures d'itinéraires actuellement envisagées ont une certaine rigidité. Dans le modèle formel que nous proposons, la structure de l'arbre est statique : nombre d'étapes à chaque niveau connu ; requêtes dans chaque étape non paramétrée, etc. Le problème posé ici consiste à définir un modèle plus souple permettant de décrire la structure et le contenu d'un itinéraire de manière paramétrée. Cela suppose que l'on puisse formaliser le passage de paramètres entre les étapes d'un itinéraire. Ce passage de paramètre peut à son tour être contraint en fonction du modèle d'itinéraire choisi.

Une telle approche peut permettre de générer des familles d'itinéraires pour un même type de métadonnées. Par exemple, générer un itinéraire pour n'importe quelle phénomène physique : pression, gravitation, etc. (« phénomène physique » serait une classe de l'ontologie dont « pression » et « gravitation » seraient deux instances particulières).

## **4.4 Adaptation utilisateur et conception d'une station auteur d'itinéraires**

A titre d'information, nous indiquons ici un travail en cours dans le cadre du projet européen (projet Mesmuses [1]).

Qui sont les futurs concepteurs d'itinéraires : interlocuteurs privilégiés des médiathécaires (pour l'indexation des documents), des graphistes, des informaticiens, etc. le concepteur d'itinéraire doit disposer à la fois d'un outil intégré qui lui facilite la création d'itinéraires mais aussi sa gestion au sens large : stockage, partage, modification, test et publication. Or un élément important de complexité dans la mise en œuvre de telles plateformes vient de la volonté de plus en plus fréquente d'introduire de l'adaptation utilisateur. Ce qui pose de nombreuses questions, tant du côté de la cognition, de l'apprentissage, etc (qu'est ce qui différencie un itinéraire pour enfant d'un itinéraire pour adulte ? Y a-t-il des structures argumentatives identifiées et différentes ?) Dans quelle mesure le ciblage d'un public est il ou non corrélé avec les niveaux de spécialisation des thesauri existant, eux-mêmes représentés dans l'ontologie ?

Or il est vraisemblable que tous ces éléments ne sont pas totalement décorrélés, d'où une volonté de factoriser ces éléments de composition d'un itinéraire. Dans ce cas, la création et l'édition font effectivement partie des tâches les plus ardue tant le concepteur est amené à gérer une combinatoire importante entre l'ontologie du domaine sur laquelle il se base, la structure de son itinéraire et les éléments de spécialisation vis-à-vis des utilisateurs. La plateforme de création d'itinéraires doit donc être à la fois un outil ouvert, suffisamment riche pour ne pas frustrer le concepteur dans la réalisation de son « œuvre », mais aussi suffisamment contraint pour qu'il ne s'y perde pas.

La version actuelle du prototype est actuellement en démonstration sur le site [démon](#)  
[3](#).

## **5. Retour sur la composition des itinéraires: le parcours comme modèle alternatif**

L'architecture de référence, décrite [plus haut](#), suppose que le document, quel qu'il soit, est toujours l'unité élémentaire d'indexation. En conséquence même une étape ne se projetant que sur un seul objet peut être reliée, par cette indexation, à un tas hétérogène de ressources du web. La section 5.1 réouvre cette question. Dans la section 5.2 nous revisitons une autre question, celle de savoir sur quoi se projette une étape, sur quelle "unité sémantique" ? La section 5.3 décrit le principe d'une nouvelle architecture logique, dont la section 5.4 montre les conséquences bénéfiques possibles tant pour le concepteur que l'utilisateur d'itinéraires générés par des parcours prédéfinis par un va-et-vient entre documents (illustrations de propos) et unités sémantiques (sujets ou intentions de propos).

### **5.1 Les documents incorporent des itinéraires par défaut**

La liberté de lecture dans les deux dimensions d'une page et dans la "troisième dimension" que constituent les hyperliens entre pages est très souvent, sinon réduite, du moins "encadrée" par le guide documentaire traditionnel que constituent la "tables

des matières" (TdM) automatiquement générée dans la conversion vers le format HTML de tout document dont le format source obéit à un "plan" (cf. les enregistrements sous format HTML des documents de source Word ou Powerpoint). Cette TdM peut générer soit une page de garde soit un cadre - un "frame" - qui reste présent, en général en marge gauche, comme un itinéraire - prédéfini et statique - de lecture du document.

L'art des concepteurs de sites documentaires sur le web fait un large usage de ces structures très suggestives d'itinéraires que sont les listes à plusieurs niveaux - parfois appelées "plans de site", les tableaux et même les figures interactives parmi lesquelles les cartes tiennent une place de choix.

Aussi les réflexions qui viennent à l'esprit quand on pense à l'indexation automatique, évoquée [plus haut](#) dans l'architecture de référence, sont naturellement:

- doit-on indexer des sites, des grappes de pages fortement liées, des pages entières ou les composants élémentaires classiques que sont les paragraphes, les tables et les figures/images s'ils sont individuellement accessibles par URL ou URL#cible ?
- doit-on récupérer en même temps ou séparément les éléments semi-structurant que sont les (sous-)titres, les TdM, les cadres ou autres "commutateurs" ?

La réponse à la première question est le plus souvent évasive, tant il est vrai que la logique de composition "maître-détail" ne peut jamais laisser prédire à quel "niveau de détail" une ressource du web sera déclarée d'intérêt pour une communauté donnée.

Ne pas répondre à la deuxième question est ambigu : n'y aurait-il dans le plan d'un ensemble documentaire cohérent aucune trace ni de l'ontologie implicite ni du genre argumentatif qu'il utilise - et que la communauté pourrait avoir intérêt à reconnaître ?

Une voie de recherche intéressante serait de récupérer automatiquement ces bouts d'itinéraires inscrits dans la construction hyper-textuelle des ressources du web pour enrichir, s'ils sont compatibles, les itinéraires préétablis par la communauté.

## **5.2 Quelles correspondances entre structures d'itinéraires et ontologie ?**

L'un des aspects les plus ardues de notre recherche concerne la définition de correspondances entre une structure d'itinéraire et l'ontologie qu'elle utilise. Dans les sections précédentes nous avons simplement supposé qu'à une étape était associée un "contenu" et que ce contenu pouvait provenir d'une "projection" sur une partie du graphe ontologique et, par transitivité, sur l'ensemble des ressources du web indexées par les objets ainsi "éclairés".

Si l'ontologie est gérée par une base de méta-données semi-structurées (RDF/RDFS), est-ce que le contenu sémantique d'une étape peut être défini par une requête sur cette base (RQL/RSSDB) ? Autrement dit est-ce qu'une étape peut n'avoir comme intention descriptive (ou argumentative) qu'un objet, une collection d'objets, une classe d'objets ou plus généralement une "vue" (c'est à dire le résultat d'une requête logiquement bien formée) sur l'ontologie ?

Cette question nous a conduits à faire l'état de l'art sur les vues dans le cadre des modèles à objets [24] ou celui des modèles de données semi-structurées [25, 26]. Aucune théorie complète satisfaisante ne semblant encore proposée, nous avons choisi de nous limiter aux cas où une étape peut être associée une intention descriptive simple: un objet, une collection d'objets ou une classe d'objets.

Il reste que l'ensemble des documents/ressources du web associées par indexation à ces objets - opération semi-automatique préalable selon l'architecture de référence décrite plus haut - risque de présenter une grappe abondante de documents candidats dans laquelle le concepteur d'itinéraire aura à faire un élagage important.

### **5.3 Définition coordonnée des itinéraires ontologiques et des itinéraires documentaires.**

Les réflexions précédentes nous ont conduits à imaginer un modèle alternatif dans lequel les itinéraires seraient construits, ni directement sur les documents-ressources du web, ni sur les objets ou collections de l'ontologie, mais sur des éléments de construction descriptive (ou argumentative, comme envisagé dans la section précédente) intermédiaires car doublement reliés à l'ontologie d'une part - par leur "objet/sujet" - et à l'ensemble documentaire d'autre part - par leurs "illustrations", cf. figure 4 ci-après.

Nous appelons "**propos**" ces éléments évidemment liés entre eux par une relation de composition "maître-détail".

Nous posons que :

- un **propos élémentaire** n'a qu'une et une seule "intention", que l'on pourrait aussi appeler son titre par analogie avec celui d'un document qui l'implémenterait ou, au moins, l'illustrerait (cf. ci-après)
- le "sujet" d'un propos est un objet de connaissance (objet "concrêt", "abstrait" - une classe - ou résultat de requête)
- l' "illustration" d'un propos est un objet documentaire (simple ou composite)
- un propos a 0 ou **1 sujet** et 0, 1 ou **n illustrations**; un propos naissant n'a qu'une intention - un titre - il se fixe ensuite un sujet dans une ontologie, il se réalise enfin en documents qui l'illustrent - successivement (versions de développement), alternativement (versions pour différents publics) ou complémentaiement.

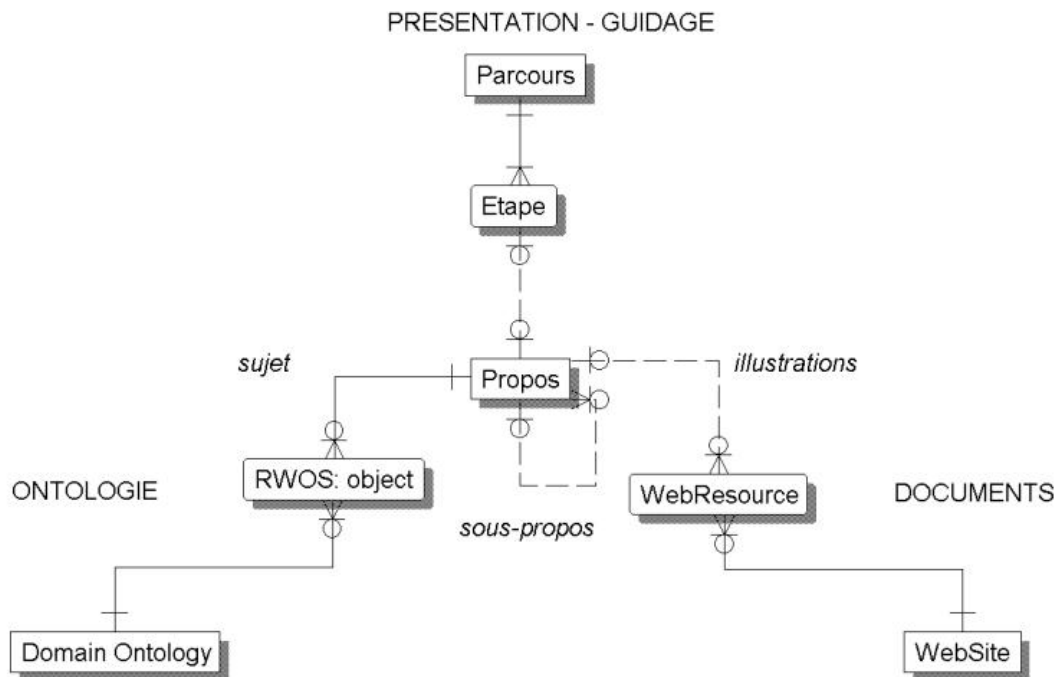


Figure 4 : modèle de parcours

Appelons alors "**parcours**" un itinéraire construit sur de tels propos, à raison d'un propos par étape. Une étape - ou "station" - dans un tel parcours est une "unité sémantique" de présentation (ou "discours") descriptive (ou argumentative) qui peut être visitée par un ou plusieurs chemins. Nous dirons qu'une "station" est sur un seul "fil" sémantique - une seule "ligne" - ou sur plusieurs. Un lecteur-visiteur garde la liberté de suivre tel ou tel trajet, comme dans tout itinéraire balisé.

Cette conception suggère fortement que les propos implicites dans les documents pourraient faire partie de la construction des parcours et donc être bien reconnus dans la phase d'indexation. Elle permet aussi une plus grande liberté pour la spécification progressive des parcours :

26. *top-down*: construire une "bonne" ontologie; développer des propos à son sujet; les assembler en discours; y reconnaître des itinéraires, voire des parcours et des fils; enfin chercher ou fabriquer les documents les plus illustratifs
27. *bottom-up*: rassembler, comme autant de bookmarks, une collection documentaire "utile" à un discours sur une ontologie; reconnaître des "unités sémantiques" induites et des transitions pour constituer un premier graphe; tenter des zoom-in pour descendre jusqu'à des propos qui ont des sujets bien identifiés dans l'ontologie (requêtes)
28. *mixte*: cf. [démon 2](#) illustrant la construction de parcours avec un simple éditeur RDF/RDFS comme Protégé-2000 étendu pour y incorporer des fonctions de

récupération automatiques de documents désignés par URL, et de génération automatique de site de parcours.

## **5.4 Intérêt de cette nouvelle approche de la définition des itinéraires**

### **5.4.1 Généricité, cohérence structurelle et sémantique des itinéraires**

Les problèmes pratiques auxquels permettrait de répondre cette étude sont les suivants :

- Peut-on, à partir d'un itinéraire existant sur la pression (métadonnée instance de la classe « phénomène physique ») générer un itinéraire sur la gravitation, métadonnée instance de la même classe « phénomène physique » ? Dans quelle mesure la structure de l'itinéraire initial peut elle être récupérée avec ou sans modification (on peut sans risque supposer que des différences structurelles risquent d'apparaître : par exemple, le nombre de savants ayant travaillé sur chacun des phénomènes et donnant lieu à une étape peut être différent).
- Peut-on à partir de métadonnées sélectionnées par un utilisateur, construire dynamiquement un itinéraire explorant ces métadonnées ? On peut imaginer différents types de réponse sur la base de « mélange » d'itinéraires existants ou de création totalement nouvelle d'itinéraires.

Cette liste n'est évidemment pas exhaustive. Nous aurions pu parler de la cohérence sémantique d'un itinéraire (sa capacité à traiter complètement d'un ensemble de concepts identifié), sa cohérence structurelle (les contraintes exprimées dans la structure de l'itinéraire autorise-t-elle des parcours), etc..

### **5.4.2 Itinéraires et métaphore routière**

On a vu [plus haut](#) ce que nous voulions conserver de la métaphore cartographique: un graphe où les noeuds sont les seules places où il est utile de passer du temps (de lecture, de visite), les arcs sont des transitions possibles (sous certaines conditions de profil utilisateur ou de prérequis dans l'ordre des visites) quasiment instantanée (sauf si l'utilisateur doit faire un choix). Un modèle de graphe en réseau de Pétri pourraient être une bonne traduction de notre formalisme.

Avec un tel formalisme, la métaphore du réseau routier peut être revisitée pour permettre quelques opérations très communes dans la navigation quand la décomposition hiérarchique du réseau est possible par remplacement d'un noeud par un sous-réseau connexe (sous-stations détaillées liées entre elles) respectant les liaisons du noeud remplacé (station non détaillée) :

- navigation à échelle fixe ("panoramique"): tous les noeuds sont décomposés d'un même nombre de niveaux de détail et le trajet se fait de proche en proche
- navigation à échelle variable ("pan and zoom in or out"): les noeuds sont décomposés/recomposés d'un ou plusieurs niveaux de détail au fur et à mesure du trajet

Projetés sur les objets ou collections de l'ontologie par l'application

"sujet(propos\étape)" les trajets à échelle variable peuvent correspondre à des panoramiques et des zooms "sémantiques", car en décomposant les propos on effectue bien souvent des opérations qui dans l'ontologie sont des décompositions de classe ou collection en sous-classes ou sous-collections. Réciproquement le repérage dans l'ontologie de panoramiques ou zooms sémantiques peut inspirer directement la spécification de propos intéressants pour les parcours.

Cette question du "zoom sémantique" à fait l'objet d'un [rapport interne](#) [27], demandé par le projet Mesmusés [1], pour baliser le chemin de nouvelles recherches.

## 6. Conclusion et perspectives

Consacré à l'étude du paradigme d'itinéraires dans le cadre du web communautaire, le projet e-parcours a permis d'amorcer une réflexion de fond entre ses partenaires parisiens et bretons. Si certaines pistes énoncées en début de projet se sont révélées parfois infructueuses (vue ou version sur le modèle RDF, trop proche des travaux existant sur les SGBD orientés objet), les réunions de travail, ainsi que les stages ou premières années de thèse financés par e-parcours, ont permis :

- La connaissance d'un domaine
- L'identification de problématiques de recherche
- La réalisation de prototypes illustrant notre problématique

Ils ont également permis de :

- Conforter le positionnement recherche pris par l'ENST Bretagne dans le projet de recherche européen [MESMUSES](#).
- Faire émerger, au sein même des Ecoles, des collaborations qui n'étaient pas initialement prévues dans e-parcours (par exemple le travail sur l'utilisation de la RST pour la structuration des itinéraires fait intervenir le département Langues et Culture de l'ENST Bretagne via Janet Ormrod)
- Initier des contacts avec [Mondeca](#) l'acteur français le plus actif dans les projets Topic Map
- Nous inscrire dans réseau thématique [Web Sémantique](#) créé tout récemment par le CNRS
- Imaginer une nouvelle utilisation navigationnelle de l'univers 3D du projet [VREng](#) de l'ENST
- Proposer notre compétence dans la préparation d'offre de projet sur le "Knowledge Management" aux réseaux nationaux de la recherche.

Sur le plan des perspectives, e-parcours aura des conséquences à la fois en termes de publication et de réalisation :

- Concernant les publications, deux au moins des trois thèses en cours devraient donner lieu à des publications dans le courant de l'année 2002.
  - Concernant les prototypes, il est prévu de compléter le "prototype itinéraires" (démon 3) en enrichissant les travaux sur les structures d'itinéraires et le "prototype parcours" (démon 2) en activant toutes les propriétés descriptives DublinCore, étapes et propos.
-

## 7. Annexes :

- [d mo 0](#) : diaporama lin aire g n r  par MS Powerpoint = 1 ressource du web = 1 hyperdocument
- [d mo 1](#) : itin raire purement documentaires d finis par organisation de bookmarks/signets = diaporama en r seau
- [d mo 2](#) : itin raire en suivant des propos ayant sujet ontologique et illustrations documentaires = parcours en r seau
- [d mo 3](#) : itin raire avec gestion automatique de pr requis

## Bibliographie

- [1] "MEtaphor for Science MUSEum", IST project Mesmuses IST-2000-26074, Annex 1, final version, September 2000, <http://cweb.inria.fr/Projects/Mesmuses/>
- [2] "On Storing Voluminous RDF Descriptions: The case of Web Portal Catalogs", S. Alexaki, V. Christophides, G. Karvounarakis, D. Plexousakis, [4th International Workshop on the the Web and Databases \(WebDB'01\)](#) - in conjunction with [ACM SIGMOD/PODS](#), Santa Barbara, CA, May 24-25, 2001
- [3] "Technologie interactive dans les mus es fran ais : sc nographie interactive et dispositifs p dagogiques", Bernadette Goldstein, La lettre de l'OCIM, suppl ment du n  57, 1998, pages 16-18.
- [4] « La recherche et l' volution des nouvelles technologies dans les mus es : de la base de donn es   l'interactif grand public », *Bernadette Goldstein*, p229-235, Mus e et recherche, OCIM, ISBN 2-11-089008-8, 1993
- [5] "Qui visite les mus es des sciences?", La lettre de l'OCIM, n  55, janvier-f vrier 1998.
- [6] "L'homme et les g nes", A. Kahn, CSI, version du 21/9/2000.
- [7] "Etude comparative technique et p dagogique des plates-formes pour la formation ouverte et   distance", Minist re de l' ducation nationale, de la recherche et de la technologie, Direction de la Technologie, Sous-direction des technologies  ducatives, des technologies de l'information et de la communication (DT/SDTETIC), septembre 1999.
- [8] "Rhetorical Structure Theory: Toward a functional theory of text organization", W.C. Mann, S.A. Thompson, Text, 1988, 8 (3).
- [9] "What the Semantic Web can represent", Tim Berners-Lee, W3C, <http://www.w3.org/DesignIssues/RDFnot.html>, 1998  
Mann, W.C., & Thompson, S.A. 1988.
- [10] D. Fensel et al.: *OIL in a nutshell* In: Knowledge Acquisition, Modeling, and Management, Proceedings of the European Knowledge Acquisition Conference (EKAW-2000), R. Dieng et al. (eds.), Lecture Notes in Artificial Intelligence, LNAI, Springer-Verlag, October 2000.

- [11] M.C.A. Klein et al.: *The Relation between Ontologies and Schema-Languages: Translating OIL-Specifications to XML-Schema* In: Proceedings of the Workshop on Applications of Ontologies and Problem-solving Methods, 14th European Conference on Artificial Intelligence ECAI-00, Berlin, Germany August 20-25, 2000.
- [12] N. F. Noy, M. Sintek, S. Decker, M. Crubezy, R. W. Ferguson, & M. A. Musen. Creating Semantic Web Contents with Protege-2000. IEEE Intelligent Systems 16(2):60-71, 2001.
- [13] [\*Resource Description Framework \(RDF\) Model and Syntax Specification\*](#)  
W3C Recommendation 22 February 1999
- [14] [\*Resource Description Framework \(RDF\) Schema Specification 1.0\*](#)  
W3C Candidate Recommendation 27 March 2000 [topic maps]
- [15] Dublin Core Metadata (CR Meeting Helsinki), <http://www.synec-doc.be/abd-bvd/misc/metadata.html>
- [16] On-To-Knowledge: Content-driven Knowledge-Management through Evolving Ontologies, <http://www.ontoknowledge.org/>
- [17] ICS-Forth, [The RDF Schema Specific DataBase \(RSSDB\)](#)
- [18] ICS-Forth, [The RDF Query Language \(RQL\)](#)
- [19] "Platform for Internet Content Selection" (PICS), <http://www.w3.org/PICS>
- [20] "Open Directory Project", <http://dmoz.org/about.html>
- [21] "TopicMaps.org" home page, <http://www.topicmaps.org/>
- [22] "OntoBroker Project", Université de Karlsruhe, [http://ontobroker.aifb.uni-karlsruhe.de/index\\_ob.html](http://ontobroker.aifb.uni-karlsruhe.de/index_ob.html)
- [23] "Web communautaire", C-Web, INRIA, <http://cweb.inria.fr/>
- [24] "Mécanismes de vue dans les SGBD orienté objet", Philippe Picouet, [document de travail](#)
- [25] "Quilt et les langages de requête XML", Talel Abdessalem, [document de travail](#)
- [26] "Définition de schémas externes pour le C-Web", Khaled Jouini, [rapport de stage de DEA](#) sous la responsabilité de Talel Abdessalem
- [27] "A Framework for Dynamic Exploration in Semantic Portals", Jean-Marc Saglio et Tuan Anh Ta, [rapport interne](#) de recherche