

Un méta-modèle basé sur les Topic Maps pour la structuration et la recherche d'information



Manuel Zacklad, Jean Caussanel,
Jean-Pierre Cahier

Laboratoire Tech-CICO – UTT

Quel Web Sémantique ?

- **Web Sémantique** est à la mode... mais dans la perspective d'une automatisation accrue (cf. les agents intelligents)
- **Sémantique** en un autre sens : utilisateurs déposent de l'information d'une manière qui fasse sens pour eux, construisent ensemble ce sens, s'imposent des « normes sémantiques » -> **Web sémantiquement normé**
- Différence entre :
 - **Un WEB computationnellement sémantique (sémantique formelle)**
 - **Un WEB cognitivement sémantique (significations partagées)**



Deux pôles du Web Sémantique

- Web **computationnellement sémantique**
 - Priorité n°1 : **l'inférence**
 - Recours à des agents **automatiques**
 - Langages **logiques** ou orientés objet (Ontolingua, DefOnto, KA2)
- Web **cognitivement sémantique**
 - Priorité n°1 : une indexation permettant de guider la **recherche** et la **navigation** d'un acteur humain
 - Faciliter **l'indexation partiellement manuelle** de contenus **évolutifs**
 - Langages de description **normalisés** mais suffisamment **souples et ergonomiques**

Critères de caractérisation du WEB cognitivement sémantique

- Ces critères doivent inclure :
 - Les efforts de **conception initiale** de la base, un travail d'IC dès que les contenus sont « **riches** »
 - Les **maintenances au « fil de l'eau »** au fur et à mesure de l'évolution des connaissances et de l'apparition de contenus inédits
 - La facilité **d'évaluation de la pertinence** du résultat des requêtes
- *Concernant l'orientation des recherches : ne pas faire les mêmes erreurs qu'avec les Systèmes Experts et **équilibrer les recherches** entre le pôle computationnel et le pôle cognitif*

Critères de différenciation entre les deux pôles du Web Sémantique

	Web computationnellement sémantique	Web cognitivement sémantique
Coût initial de la modélisation	Très élevé	Moyen
Mise à jour du modèle	Très élevé	Moyen
Lecture du modèle par un internaute quelconque	Quasi impossible	Moyennement difficile
Validité logique des inférences	Garantie	« Inférences incomplètes » nécessitant des interactions avec l'utilisateur pour être supervisées et interprétées
Largeur de couverture des documents par les modèles	Faible à cause du coût élevé de formalisation	Plus importante
Richesse de l'interrogation	Faible concernant la quantité de documents prise en compte, élevée concernant la précision	Elevée concernant la quantité, faible concernant la précision
Richesse d'interprétation des informations	Faible – informations ponctuelles en réponse à des requêtes ciblées	Elevée – informations riches parfois en partie inattendues



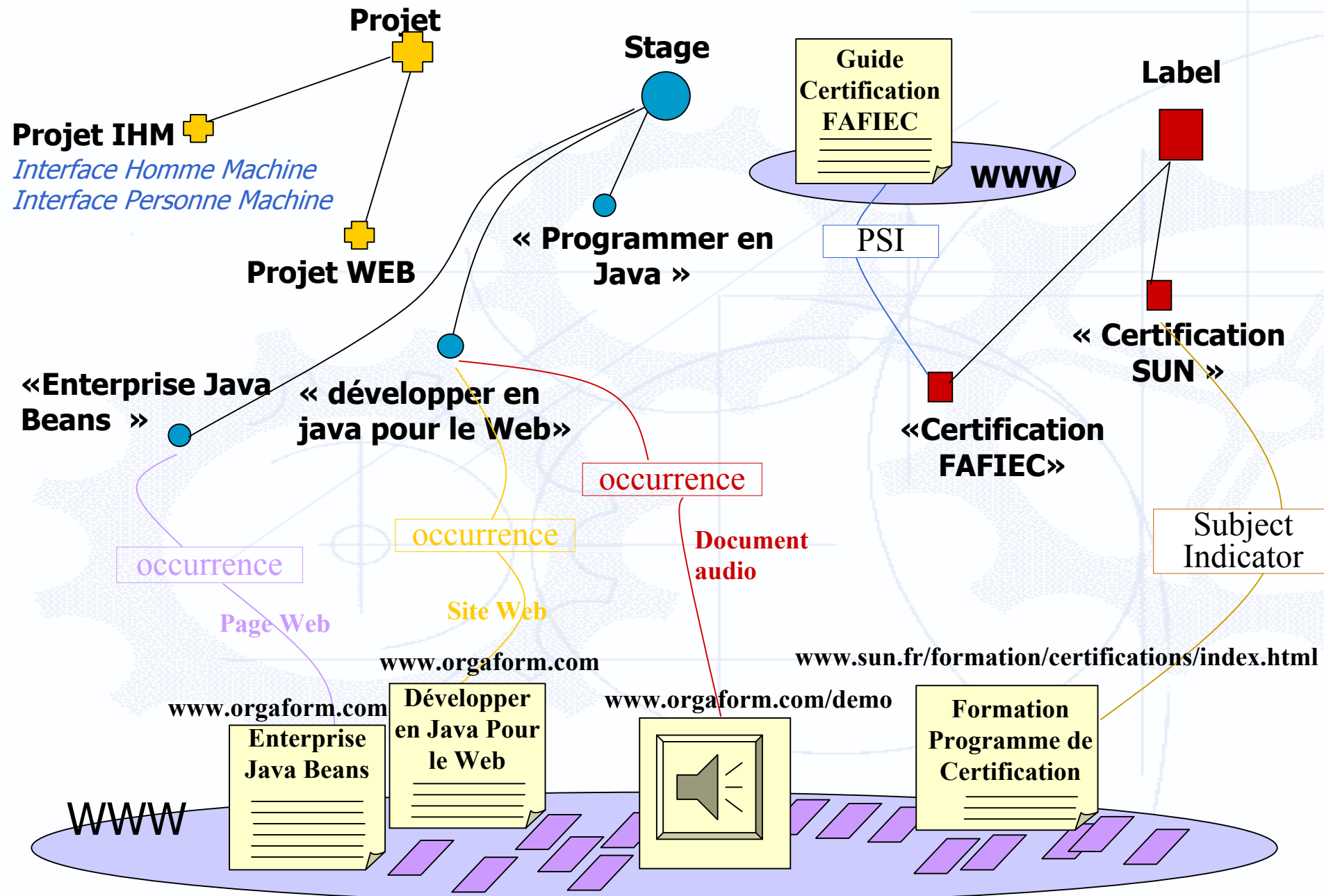
Finalité de la recherche

- Recherche d'un modèle conceptuel adapté à la construction de « place de marché à base de connaissances » (KBM : Knowledge Based MarketPlace)
- Classement d'entités « riches en signification » souvent représentées par des documents composites : catalogue actif
- Volonté d'accès à ces entités par leurs caractéristiques « standards » mais aussi à partir d'une représentation plus heuristique de leur contenu
- Coopération sur la définition du contenu traitée à travers la gestion de points de vue multiples
- Évolution assez rapide des sources représentant les entités et de la description de leur contenu



Statut des Topic Maps et d'HyperTopic

- Les Topic Maps sont un langage de représentation des connaissances de type « réseau sémantique » permettant de représenter des ontologies peu « formelles » (normalisation faible de la signification)
- Le langage Topic Maps et ses composants (thèmes, occurrences, ressources, scope...) définit aussi une ontologie de « haut niveau » d'inspiration documentaire pour décrire « le monde »
- Elles se sont avérées trop peu structurées pour l'acquisition et la gestion des connaissances des « catalogues actifs » des KBM
- En utilisant les Topic Maps comme un langage de représentation des connaissances nous définissons HyperTopic comme une « ontologie de tâche » (ou « ontologie d'usage ») pour la structuration et la recherche d'information
- Ontologie de tâche : vocabulaire lié à une tâche ou à une activité générique spécialisant une ontologie de haut niveau (tutoriel ECAI Maedche, Staab 2002)



Projet

Stage

Guide
Certification
FAFIEC

Label

Projet IHM
Interface Homme Machine
Interface Personne Machine

« Programmer en
Java »

WWW

Projet WEB

PSI

«Enterprise Java
Beans »

« développer en
java pour le Web »

« Certification
SUN »

«Certification
FAFIEC»

occurrence

occurrence

occurrence

Document
audio

Subject
Indicator

Page Web

Site Web

www.orgaform.com

www.sun.fr/formation/certifications/index.html

www.orgaform.com

Développer
en Java Pour
le Web

www.orgaform.com/demo

Formation
Programme de
Certification

WWW



HyperTopic : un modèle de compromis

- Compromis entre les langages structurés de type RDF et le langage pragmatique mais très ouvert des Topic Maps
- Compromis entre la représentation stéréotypée de l'entité et la représentation peu formalisée des thèmes et de leurs relations
- Compromis entre la vitesse de renouvellement élevée des ressources, et la vitesse de renouvellement plus faible des entités et du réseau de thèmes



La représentation intermédiaire de « l'entité »

- Importance d'une représentation structurée intermédiaire qui agrège des informations primaires caractéristiques du « signifié » véhiculé par les ressources
- Notion de produit ou de service singulier décrit « matériellement » par de multiples ressources (stage de formation, projet, sujet de veille, acteur déterminé...)

La représentation heuristique du « contenu » par le réseau des thèmes

- Les thèmes sont des expressions : « technologies réseaux pour le multimédia » ou « terminaux pour la sphère professionnelle » (cf. catégories « à la Yahoo »)
- Pas de définition intentionnelle précise du thème par des attributs
- Normalisation de la signification selon un principe de sémantique différentielle dont la validité est « locale » (dans une branche d'un arbre)
- Ontologie de domaine sémantiquement « impures » organisées en plusieurs hiérarchie « type-de » correspondant à des points-de-vue sur l'entité
- Modélisation heuristique des entités qui s'appuie en particulier sur le contenu des documents qui les décrivent
- Seule solution pratique trouvée pour modéliser un domaine de connaissance innovant (ontologies de haut niveau ou ontologie normalisée du domaine s'avérant peu utiles)
- Mode de recherche et de navigation spécifique et différente de celle utilisée pour les entités



Le rattachement des entités aux thèmes

- Dans un modèle de type « induction de concept » le rattachement de l'entité aux thèmes (selon un même point de vue) devrait se faire de manière exclusive à l'un des frères
- Dans HyperTopic le rattachement de l'entité aux thèmes est basée sur un principe de « contingence sémantique » :
 - C'est bien la possession de traits sémantiques exclusifs qui justifie la séparation entre les fils
 - Mais la même entité peut « accidentellement » posséder ces traits sémantiques



Exemple de contingence sémantique dans le rattachement de l'entité aux thèmes

- Pour une automobile les caractéristiques, « tout-terrain », « sportive », « routière », sont sémantiquement exclusives
- Mais pour un modèle donné on pourra trouver un modèle mixte tout-terrain et routier
- Justifications de la contingence sémantique :
 - Compromis entre deux traits sémantiques
 - Évolution dans le temps (traits opposés mais pas au même moment)
 - Caractéristique pertinente pour certaines parties de l'entité mais pas pour d'autres
 - Expression d'une spécification souhaitée mais non encore réalisée
 - etc....
- Création probable ultérieurement d'une nouvelle catégorie exclusive des précédentes qui remettra pour un temps « de l'ordre »

Conclusion

- Une fois dégagée l'importance de la séparation entre la définition « standard de l'entité » et la représentation heuristique des thèmes, progresser sur le réseau de thèmes :
 - Comment est-il maintenu collectivement et comment s'assurer de sa cohérence ?
 - Quelles aides à la navigation peuvent être offertes : visualisation, moteur de recherche... ?
 - Comment peut-on évaluer la pertinence de sa structuration par rapport à une recherche ?
 - Comment peut-on normaliser progressivement sa sémantique pour aller vers de l'inférence ?
 - Comment fournir une aide pour le rattachement de nouvelles entités au réseau sémantique ?