

Contexte, application : Adaptation contenu d'une page web en fonction du profil d'un utilisateur afin d'**augmenter l'utilisabilité** suivant un profil.

Démarche :

1. Récupérer, repérer des éléments d'une page web afin de **requalifier des nœuds de l'arbre** en leur associant des types définis à priori qui nous intéressent.
2. **Acquisition, représentation et stockage** des souhaits de personnalisation de l'utilisateur. On souhaite notamment étudier diverses approches dont une approche **d'acquisition et d'apprentissage à partir d'exemples**

Problématiques :

- Difficultés d'une analyse syntaxique d'une page écrite en html, c'est une format de **données semi-structurées**. Les documents ne possèdent pas de structure formelle.

Piste : Algorithme dans le domaine de la RI (qlq article lu mais inconvénients extraction de la structure pour qu'elle soit comprise par un agent et non un humain), Algorithme d'extraction de contenu basé sur des indicateurs visuels (ce qui est compris par un humain).

- **Chaque page web définit son modèle** (charte graphique différentes pour chaque site).

Piste : Transformation des modèles sources vers un modèle intermédiaire définissant une page web comme la norme HTML5, c'est à dire avec des **éléments de navigation, des articles** ect...

- Associer des types définis à priori aux structures extraites

Piste : **appariement de graphe**

- **Transformation de modèle** :

Quelles données sont manipulées en entrée ?

Spécifier, exprimer des **contraintes**. Comment respecter **la structure logique** d'une page web pour ne pas générer de conflit après des transformations. Préservation d'aspect. Comment prendre en compte le caractère interactif d'une page internet ???

Piste : en réification des structures sémantiques de haut niveau pour exprimer des contraintes ?.

*Document web représente des données hétérogènes (finalement plutôt problème lié à la recherche d'information ??? on s'intéresse seulement à la structure !???)*

Comment analyser la structure d'une page web de manière la plus automatique et générique possible. On veut récupérer **des structures de haut niveau** qui nous intéressent et lui associer des **types définis à priori**.

Piste : comment analyser la norme HTML 5 la structure d'une page web ?

