

Apprentissage d'une hiérarchie de concepts pour la conception de modèles de domaine d'hypermédias

Hermine Njike Fotzo, Thierry Artières, Patrick Gallinari, Julien Blanchard, Guillaume Letellier
LIP6, Université Paris 6
8 rue du capitaine Scott, 75015, Paris, France
{Prénom.Nom}@lip6.fr

Résumé. Nous décrivons comment apprendre automatiquement une hiérarchie de concepts à partir d'une collection de documents. Les concepts, identifiés par des ensembles de mots-clés, sont organisés en une hiérarchie de type spécialisation/généralisation. Cette hiérarchie peut être utilisée pour construire un modèle de domaine pour des collections de documents hypermédias. Nous proposons des idées sur la façon de construire des modèles d'utilisateurs à partir de tels modèles de domaine. Les modèles d'utilisateurs et de domaine peuvent être visualisés à l'aide d'outils efficaces comme les *Treemaps*.

1 Introduction

Un hypermédia adaptatif est personnalisé, dynamiquement, en fonction de l'utilisateur. La personnalisation peut consister en une adaptation du contenu de l'hypermédia ou à des aides à la navigation en ajoutant/enlevant des liens (Brusilovsky 1996). La personnalisation repose sur un modèle de l'utilisateur. Il n'existe pas aujourd'hui de consensus sur la définition d'un modèle d'utilisateur, souvent défini de façon ad-hoc, mis à part pour les hypermédias éducatifs et les systèmes tutoriaux (Da Silva 1998, De Bra 2003, Henze 1999). Ces systèmes utilisent un modèle de domaine, conçu manuellement, à partir duquel on définit les modèles d'utilisateurs. Un modèle de domaine est un graphe des concepts abordés dans l'hypermédia, caractérisant l'ensemble des connaissances accessibles dans l'hypermédia. On utilise alors des modèles d'utilisateurs du type *Overlay user models*, qui partagent la même représentation que le modèle de domaine. Ce sont des vecteurs d'attributs (un pour chaque concept) qui représentent une mesure d'intérêt ou de connaissance de l'utilisateur dans les concepts (Da Silva 1998, De Bra 2003, Kavcic 2000). Les modèles sont mis à jour à partir de la navigation de l'utilisateur. On peut également faire de l'inférence dans ces modèles en utilisant le formalisme des réseaux bayésiens (Da Silva 1998, Henze 1999).

Dans ce papier, nous nous intéressons à la définition automatique d'un modèle de domaine pour un hypermédia quelconque, à partir de son contenu. Il s'agit d'une étape préliminaire pour construire des versions adaptatives de systèmes hypermédias quelconques. Nous présentons ici une approche qui permet d'apprendre automatiquement une hiérarchie des concepts abordés dans un corpus de documents textuels, à partir du contenu de ceux-ci. Cette hiérarchie traduit une relation de spécialisation/généralisation entre les concepts. Comme nous le montrons, cette approche fournit une vue hiérarchique alternative du site web. Cette représentation du contenu thématique d'un site web permet de définir des modèles utilisateurs plus appropriés, en utilisant par exemple le formalisme des modèles bayésiens pour l'inférence. En outre, cette représentation associée à un outil de visualisation (tel que les *Treemaps*) permet la maintenance et/ou l'analyse des modèles d'utilisateurs.