
Fermer

Imprimer

Titre : Construction automatique de présentations brèves pour la recommandation de textes courts

Sujet proposé dans : M2R Informatique, Projet

Responsable(s) :

- Catherine Garbay (Catherine.Garbay@imag.fr) LIG-AMA
- François Portet (Francois.Portet@imag.fr) LIG-GETALP
- Sybille Caffiau (sybille.caffiau@imag.fr) LIG-IIHM

Mots-clés : extraction d'informations, génération de récit, texte

Durée du projet : 6 mois, possibilité de continuer en thèse

Nombre maximal d'étudiants : 1

Places disponibles : 1

Interrogation effectuée le : 29 septembre 2014, à 13 heures 09

Description

Contexte

L'objectif est de construire de manière automatique des présentations brèves permettant la mise en valeur des textes courts publiés par la société Short Edition et facilitant l'accès à ces nouvelles pour le lecteur. Ces présentations jouent le rôle des « quatrièmes de couverture » classiquement utilisés pour les romans dans le domaine éditorial. Le travail se fera en partenariat avec la société Short Edition.

Enjeux

La construction doit suivre un schéma narratif, en associant par exemple un résumé (ou extrait), quelques informations sur l'auteur et les critiques faites à son sujet ou le point de vue de l'éditeur. Le contenu et le style adopté doivent susciter l'attention, mais demeurer fidèles au contenu et à la tonalité du roman.

Il faut tenir compte également des préférences des lecteurs (certains peuvent préférer trouver un résumé plutôt qu'un extrait), et éventuellement désigner de manière explicite une classe de lecteur (âge de lecture recommandé).

La présentation reflète enfin une politique éditoriale, et peut dépendre des collections présentées.

Objectif du projet

Il existe une littérature abondante concernant (i) l'extraction d'information à partir de textes et (ii) la génération automatique de textes. Les systèmes de génération actuels ont tendance à s'appuyer sur un mode de présentation « factuel » pour transmettre des informations, sans adoption de point de vue ni de mise en perspective d'éléments particuliers dans le contenu à transmettre. Dans ce projet, la présentation doit être à la fois orientée par les exigences propres à la construction narrative, fidèle aux éléments factuels, et sensible aux préférences individuelles et au point de vue éditorial.

Au plan méthodologique, on disposera d'un corpus de textes courts, de données biographiques sur leurs auteurs, et de commentaires émis par le Comité éditorial. La construction de la présentation s'appuiera sur des modèles de tâches définis sous K-MADe (Kernel of Model for Activity Description - www.lias-lab.fr/forge/projects/kmade) et reflétant l'ensemble des schémas narratifs et des propriétés des présentations à construire. Cette modélisation reposera sur une formalisation de la connaissance à partir d'ontologie pour assurer une unité sémantique aux traitements ainsi que pour des raisons de généralité (externaliser les connaissances appartenant purement au domaine d'application).

Il faudra ensuite concevoir le système dont le rôle sera d'extraire et de sélectionner les éléments constituant la présentation et de les organiser sous forme d'un plan de récit. Celui-ci sera formalisé comme un graphe d'éléments reliés par des relations syntaxiques et sémantiques. Ce graphe enrichi pourra ensuite être transformé sous la forme d'un texte narratif via des techniques de génération du langage naturel.

La sélection des schémas narratifs, des éléments et niveaux de granularité pertinents constituera une tâche centrale.

Son objectif est de permettre l'adaptation de la construction du graphe selon le point de vue éditorial adopté ou les préférences des lecteurs. Ces décisions s'appuieront sur des connaissances du domaine d'application et sur des connaissances plus théoriques du domaine de la linguistique et de la narratologie. Des modèles hybrides statistico-relationnels comme les Réseaux Logiques de Markov seront utilisés pour apprendre et exprimer des relations syntaxiques et sémantiques à différents niveaux de granularité entre éléments de la présentation. L'évaluation des présentations ainsi construites associera des membres du Comité éditorial de l'entreprise partenaire et des participants « naïfs ».

Fermer

Imprimer