

Narrative Summarization

Claude Cunha¹ Guillaume Le Garrec¹

(1) IRIT, Université de Paris, 7 rue Thomas Mann, 75013 Paris, France

{claude.cunha, guillaume.lgarrec}@etu.univ-paris-diderot.fr

MOTS-CLÉS : RÉSUMÉ AUTOMATIQUE DE TEXTES, RÉCITS, TRAME NARRATIVE.

RÉSUMÉ

Encore aujourd'hui, le résumé de textes narratifs reste un défi. Les différentes difficultés inhérentes à un tel procédé ont été décrites dans un grand nombre de papiers (Mani, 2004). Les aspects considérés comme les plus problématiques semblent être la construction de la ligne temporelle d'une histoire ainsi que la prise en compte des différentes intrigues et sous-intrigues se déroulant en même temps. Des travaux précédents décrivaient des représentations pour les relations temporelles dans les textes narratifs (Zhou 2006, Mani and Pustejovsky, 2004), mais aucun ne présentait une procédure complète pour le faire. Ceci est principalement dû à la difficulté de traiter les informations implicites. Par exemple, un personnage se souvenant d'un évènement ne donne pas nécessairement la date ou la période à laquelle il s'est déroulé. Il est possible que le lecteur soit supposé pouvoir situer l'évènement dans le temps en fonction du contexte dans l'histoire à résumer. Plus généralement, le problème est de réussir à créer un moyen complet pour modéliser ce qui se passe dans un texte narratif. Avec ces études en tête, nous avons essayé d'adopter une autre approche pour simplifier le processus de résumé. Ainsi, nous n'avons pas développé un algorithme pour construire le modèle d'une histoire afin d'en dériver un résumé. Au lieu de cela, nous avons essayé de créer un ensemble de critères permettant de sélectionner des phrases du texte pour constituer son résumé. Ces critères reposent sur la position des phrases dans le texte, la fréquence d'apparition des synsets représentés par les mots d'une phrase, la fréquence d'apparition des noms propres dans la phrase ainsi que la longueur d'une phrase. Nous avons conservé les phrases de moins de vingt mots, car nous avons observé que les phrases plus longues étaient des phrases de description. Nous avons également pénalisé les phrases trop courtes, qui sont peu narratives. En sélectionnant des phrases selon ces critères nous avons produit un résumé qui se voulait un rapide coup d'oeil du livre. Nous avons testé cette méthode sur les sept premiers chapitres du roman *Oliver Twist* de Charles Dickens. Le résumé obtenu était incohérent car les phrases obtenues ne formaient pas d'histoire à proprement parler. Nous avons choisi de ne pas exclure les phrases comportant des marques de dialogue car elles peuvent apporter des éléments narratifs. Cependant, les phrases retenues permettaient d'obtenir certaines informations intéressantes sur les sept premiers chapitres du livre comme le fait qu'Olivier est probablement le personnage principal de l'oeuvre ou que l'environnement dans lequel se déroule l'histoire est plutôt sombre.

Ce projet a été réalisé sous la supervision de Nicolas Ballier et Jean-Baptiste Yunès à l'Université de Paris.

REMERCIEMENTS

Nous remercions Nicolas Ballier ainsi que Jean-Baptiste Yunès qui nous ont permis d'entreprendre ce projet de recherche et qui nous ont guidé dans sa réalisation. Ce travail est l'œuvre conjointe d'étudiants de la Licence Informatique et de la Licence d'Études Anglophones de Paris Diderot. Il a été financièrement soutenu par le programme IdEx Université de Paris ANR-18-IDEX-0001.

Références

MANI, I. (2004). Narrative summarization. *Traitement Automatique des Langues*, 45(1):1-24.

ZHOU L., MELTON G.B., PARSONS S., HRIPCSAK G.. A temporal constraint structure for extracting temporal information from clinical narrative. *J Biomed Inform* 2006; 424-39.

MANI, I. AND J. PUSTEJOVSKY. 2004. Temporal Discourse Models for Narrative Structure. Proceedings of the ACL 2004 Workshop on Discourse Annotation.