

Tag Thunder : plateforme de démonstration et d'expérimentation

Jean-Marc Lecarpentier Elena Manishina Maxence Busson

Fabrice Maurel Stéphane Ferrari

Université de Normandie, UNICAEN, GREYC UMR CNRS 6072

Campus Côte de Nacre, Bd du Maréchal Juin, 14032 CAEN cedex 5, France

{firstname.lastname}@unicaen.fr

RÉSUMÉ

Dans cette démonstration, nous proposons un système qui permettrait aux utilisateurs non-voyants d'obtenir le *first glance* d'une page web. L'objectif est de réduire le temps d'accès à la structure logico-thématique de la page et de favoriser le développement de stratégies de lecture de haut niveau. Notre concept, appelé Tag Thunder, s'appuie sur une phase de segmentation de la page en zones, suivie d'une étape de représentation des zones par un mot ou groupe de mots, puis une vocalisation simultanée de ces représentants.

ABSTRACT

Tag Thunder : demonstration and experimentation platform

In this demonstration we present an architecture to allow visually impaired people to perceive the layout of a web page. We aim to speed up access to the page structure and the topics of its sections, as well as develop high level reading strategies. Our concept, named Tag Thunder, is based on 3 steps : first, the segmentation of a web page into areas, then keyword extraction for each area and finally a vocalization of the keywords.

MOTS-CLÉS : Synthèse de la Parole, Segmentation, Extraction des Mots Clés.

KEYWORDS: Speech Synthesis, Content Segmentation, Keyword Extraction.

1 Introduction et objectifs

La navigation visuelle sur internet permet d'accéder à une vue précoce et globale des pages Web ('skimming'). (Dias & Conde, 2007) montrent que dans ce cas la typographie et la disposition des éléments dans le document revêtent une importance capitale. Les usagers non-voyants n'ont que peu l'usage de cette possibilité (Ahmed *et al.*, 2012), bien qu'il existe des stratégies palliatives (Borodin *et al.*, 2010) telles que l'accélération du débit de la synthèse de la parole, le saut de titres en titres ou de liens hypertextes en liens hypertextes, la lecture des premières ou dernières phrases de tous les paragraphes, etc. Néanmoins, la différence d'efficacité, en comparaison avec la navigation visuelle, reste significative (Bigham *et al.*, 2007).

Dans la présente démonstration, nous proposons une approche qui permettra aux utilisateurs non-voyant d'obtenir une vue globale (*first glance*) d'une page web donnée. Notre représentation combine un filtrage du contenu informationnel d'une page avec la prise en compte de ses propriétés visuelles.

Notre concept, appelé Tag Thunder, s'appuie sur une phase de segmentation de la page en zones, suivie d'une étape de représentation des zones par un mot ou groupe de mots, puis une vocalisation simultanée de ces représentants.

2 Architecture

Cette partie présente chacune des étapes menant à la création d'un Tag Thunder.

Segmentation de la page : les éléments de la page HTML sont tout d'abord enrichis par leurs informations de position, taille et styles (Giguet & Lucas, 2011). Un premier filtre ne garde que les éléments visibles à l'écran. Nous utilisons ensuite l'algorithme k -moyennes pour regrouper les éléments HTML en zones. Les paramètres d'entrée de l'algorithme sont : le nombre de zones désirées, la distance de séparation (non aléatoire), le nombre d'itérations.

Extraction des mots clé : pour chacune des zones, nous sélectionnons une liste de mots (ou groupe de mots) représentant thématiquement la zone. Notre technique s'appuie sur une extraction des meilleurs n -grammes de chaque zone : pour chaque n -gramme nous calculons un TF/IDF (Sparck Jones, 1972). Le TF est la fréquence d'un n -gramme dans la zone d'où il est extrait. Le calcul de l'IDF (Inverse Document Frequency) a été réalisé sur un corpus de textes constitué de 953 551 articles sur 20 années du journal Le Monde (de 1987 à 2006). Les scores TF/IDF sont modulés à partir d'informations sur la disposition des n -grammes (position dans la zone) et leur mise en forme (taille de police, graisse, contrastes, etc) .

Vocalisation : l'objectif est de produire à l'oral les n -grammes en traduisant les propriétés visuelles des zones qu'ils représentent. Nous utilisons le synthétiseur vocal Kali (Morel & Lacheret-Dujour, 2001). Les caractéristiques de la zone sont associés aux paramètres acoustico-prosodiques du n -gramme de la manière suivante :

- identifiant de la zone : type de voix prononçant le n -gramme (avec une alternance homme/femme) ;
- position du centroïde de la zone : spatialisation stereo du n -gramme ;
- surface de la zone : volume du n -gramme ;
- complexité visuelle de la zone (calculée à partir du nombre d'éléments HTML dans la zone et de la somme des aires de ces éléments) : fréquence de répétition du n -gramme.

La vocalisation simultanée des n -grammes ainsi paramétrés produit le Tag Thunder.

3 Conclusion et perspectives

L'architecture présentée permet d'obtenir une représentation sonore de la structure logico-thématique d'une page web. Les expériences en cours devront confirmer que notre concept facilite la navigation non visuelle intra page web. Les travaux futurs consisteront à l'amélioration de plusieurs points de notre dispositif : combiner des techniques issues du TAL et de l'imagerie pour la phase de segmentation ; formaliser l'influence de la mise en forme des n -grammes sur les scores ; déterminer par une démarche expérimentale les meilleures associations entre les caractéristiques de la zone et les paramètres acoustico-prosodiques (utilisation de la spatialisation 3D).

Remerciements

Ce projet a reçu le soutien de la Région Normandie dans le cadre du CPER NUMNIE.

Références

- AHMED F., BORODIN Y., SOVIK A., ISLAM M., RAMAKRISHNAN I. & HEDGPETH T. (2012). Accessible skimming : Faster screen reading of web pages. In *25th Annual ACM Symposium on User Interface Software and Technology (UIST)*, p. 367–378.
- BIGHAM J. P., CAVENDER A. C., BRUDVIK J. T., WOBROCK J. O. & LANDER R. E. (2007). Webinsitu : A comparative analysis of blind and sighted browsing behavior. In *9th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, p. 51–58.
- BORODIN Y., BIGHAM J. P., DAUSCH G. & RAMAKRISHNAN I. (2010). More than meets the eye : A survey of screen-reader browsing strategies. In *International Cross Disciplinary Conference on Web Accessibility (W4A)*, p. 1–10.
- DIAS G. & CONDE B. (2007). Accessing the web on handheld devices for visually impaired people. In K. WEGRZYN-WOLSKA & P. SZCZEPANIAK, Eds., *Advances in Intelligent Web Mastering*, volume 43 of *Advances in Soft Computing*, p. 80–86.
- GIGUET E. & LUCAS N. (2011). *Comparative Evaluation of Focused Retrieval : 9th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2010, Vught, The Netherlands, December 13-15, 2010, Revised Selected Papers*, chapter The Book Structure Extraction Competition with the Resurgence Software for Part and Chapter Detection at Caen University, p. 128–139. Springer Berlin Heidelberg : Berlin, Heidelberg.
- MOREL M. & LACHERET-DUJOUR A. (2001). Kali, synthèse vocale à partir du texte : de la conception à la mise en oeuvre. *Traitement Automatique des Langues* 42, p. 193–221.
- SPARCK JONES K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1), 11–21.