Du *tag cloud* au *tag thunder* : vers de nouvelles stratégies de lecture orale pour non-voyants

Fabrice Maurel
Université de Caen Normandie
Campus Côte de Nacre, Boulevard du Maréchal Juin
CS 14032
14032 CAEN cedex 5
fabrice.maurel@unicaen.fr

Alexandre Beudin
Université de Caen Normandie
Campus Côte de Nacre, Boulevard du Maréchal Juin
CS 14032
14032 CAEN cedex 5
abeudin@gmail.com

Stéphane Ferrari
Université de Caen Campus Côte de Nacre, Boulevard du Maréchal Juin
CS 14032
14032 CAEN cedex 5
stephane.ferrari@unicaen.fr

Jean-Marc Lecarpentier
Université de Caen Normandie
Campus Côte de Nacre, Boulevard du Maréchal Juin
CS 14032
14032 CAEN cedex 5
Jean-marc.lecarpentier@unicaen.fr

http://dx.doi.org/10.1145/2820619.2825021

Résumé
Le Web est un puissant portail d'accès à l'information et à la connaissance. Les usagers non-voyants l'utilisent couramment en accédant aux pages Web avec des lecteurs d'écrans (JAWS, VoiceOver, Tackback…) ou des modules ajoutés aux navigateurs (FireVox ou ChromeVox). Cependant, le manque de perception rapide et globale des documents rend la navigation avec ces technologies encore beaucoup moins efficace et satisfaisante que lors d'une interaction visuelle. Nous introduisons dans cet article le concept de *tagthunder*, une version sonore et interactive des *tagclouds*, qui permet d'envisager un résumé multiniveau des contenus de la page à partir de l'analyse de sa structure logico-thématique. Nous espérons ainsi fournir l'équivalent oral d'un « premier regard » sur les pages Web qui favoriserait l'émergence de nouvelles stratégies de lecture non visuelles de haut niveau (lecture rapide ou en diagonale).

Mots clés choisis par les auteurs
Lecture rapide non visuelle, mise en page, transposition orale de nuages de mots

Mot clés de la classification ACM
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

La navigation visuelle sur internet permet d'accéder à une vue précoce et globale des pages Web (stratégie dite d'écrémage ou *skimming*) puis de procéder à une recherche active et rapide d'informations spécifiques (stratégie dite de balayage ou *scanning*). Dans les deux cas la typographie et la disposition des éléments dans le document revêtent une importance capitale [1] (titres, contenus proéminents, phrases spécialement positionnées). Les usagers non-voyants n'ont que peu l'usage de cette possibilité [2] (voir Figure 1), bien qu'ils aient développé des stratégies palliatives [3] 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. Néanmoins, la différence d'efficacité, en comparaison avec la navigation visuelle, reste significative [4]. Nous nous intéressons dans ce travail à la conception de stratégies non linéaires de *skimming* et de *scanning* adaptées à une navigation non visuelle afin d'améliorer les possibilités d'accès à l'information pour les usagers non-voyants (lecture rapide ou en diagonale).

L'idée que nous développons est de s'appuyer sur une transposition à l'oral du concept, très présent sur le Web, de représentation visuelle et interactive de mots-clefs : le nuage de mot (ou *tag cloud* – voir Figure 2). La construction d'une telle version audio, que nous dénommons « tonnerre de mots » (ou *tag thunder*), devra permettre d'appréhender globalement et précocement la structure logico-thématique de la page Web et favoriser ainsi l'émergence de stratégies de *scanning* non visuelles innovantes.

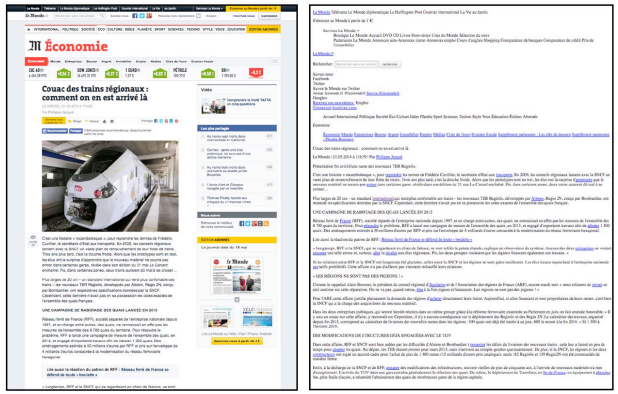

Figure 1. Comparaison entre le visuel d'une page web et sa perception par un non-voyant avec les lecteurs d'écran

Figure 2. Exemple de *tag cloud* classique

Contexte scientifique

Des études proposent de s'appuyer sur des techniques de résumé de texte pour offrir aux non-voyants la possibilité de stratégies de *skimming* [5]. Mais [1] critique le fondement de cette approche par une évaluation exhaustive des textes non structurés ; l'analyse démontre des limites pour la perception rapide des contenus de pages Web. De notre point de vue, ce qui est pointé ici est le rôle important de la typo-disposition pour une bonne appréhension des contenus et la mise en place d'une boucle perception/action efficace. D'autres propositions ont voulu intégrer cette observation dans des solutions utilisant ou combinant différentes modalités d'accès au texte et à sa mise en forme (synthèse de la parole, retours tactiles ou vibrotactiles). Nous citons ci-après celles qui ont principalement conduit notre réflexion.

Un modèle d'oralisation par reformulation est proposé par [6] pour améliorer la sensibilité des systèmes de synthèse de la parole à partir de textes (TTS) à certains phénomènes syntaxiques, typographiques ou dispositionnels (relevant de ce que les auteurs appellent l'Architecture Textuelle). Cette approche permet une amélioration sensible de la mémorisation et de la compréhension lors de l'oralisation de documents fortement structurés. Malgré tout, la charge cognitive reste moins bien gérée que lors d'une lecture visuelle de ces mêmes documents [7].

Le projet AcceSS (Acessibility through Simplification & Summarization – [8]) offre deux fonctionnalités. La première, la simplification, permet de retirer le contenu jugé moins important des pages web. La deuxième, le résumé, cherche à donner un aperçu du contenu à l'utilisateur non-voyant similaire à ce qu'un voyant perçoit. Pour ce faire, une méthode de marquage génère un ensemble de repères placés sur les éléments du document et une nouvelle page est créée pour chacune de ces sections. Une page « *Guide Dog* » sert de point central de navigation et de sommaire. Un gain de temps est observé pour accéder à l'information et la navigation améliorée. Cependant la méthode de *pattern matching* pour identifier les sections de la page est encore faillible ; de plus il n'y a pas de simplification au niveau du texte, donc une difficulté toujours grande dans le cas de contenus textuels importants.

SeEBrowser (Semantically Enhanced Browser) est un navigateur web audio spécialisé pour non-voyants [9]. Il propose une navigation par raccourcis grâce à des annotations sémantiques formant les nœuds (*browsing shortcuts* – BSs) d'un arbre de navigation ; l'utilisation de cette hiérarchie évite la surcharge d'information. Le gain de temps est statistiquement significatif pour des tâches de recherche d'information (*scanning*) mais toujours pas comparable à un accès visuel.

HearSay est un navigateur web non-visuel multimodal actuellement à sa troisième version [10]. Il supporte plusieurs sorties (audio, écran, braille), entrées (reconnaissance vocale, clavier, tactile) et fonctionnalités : un module de segmentation permet de naviguer entre les objets textuels sémantiquement reliés ; un système d'annotation gère l'ajout de texte alternatif aux images et autres contenus ; le module « Dynamo » repère les changements entre les pages ; un analyseur de contexte détecte le contenu principal et identifie les informations pertinentes. HearSay constitue une avancée significative au-dessus des meilleurs lecteurs d'écran : réduction du temps requis pour trouver le contenu principal, capacité à localiser

les changements dynamiques et éviter de lire les informations répétées, amélioration de la qualité de navigation. Cependant, l'accès à la structure n'est pas complet car il permet d'atteindre rapidement le contenu principal mais l'accès reste séquentiel, donc parfois long et laborieux, pour les autres éléments.

Dans notre approche, nous proposons une autre solution orale, économique et non-intrusive, basée sur (1) une extraction itérative de mots-clefs conditionnée par la structure logique de la page Web, et (2) la construction de *tag thunders*, versions audio et multicanales du concept de *tag cloud*.

Solution proposée

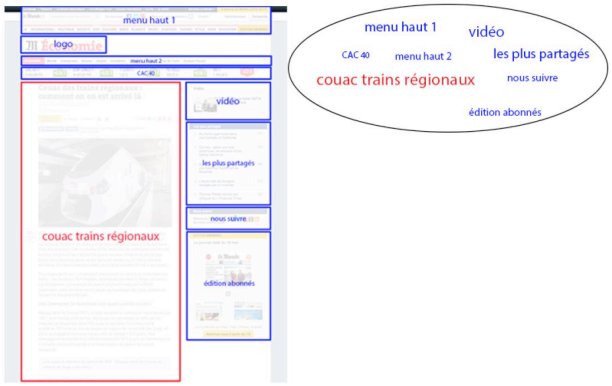

Figure 3. Division d'une page web d'article de presse et formation d'un *tag cloud*

Considérons visuellement une page Web. Nous pouvons la définir comme un ensemble de blocs. Chacun est identifiable par un ensemble de mots représentants son type (métadonnées) ou son contenu (mots-clés). Si l'on efface les autres éléments de la page, ces « tags » semblent inscrits spatialement dans la page selon (1) la position occupée par le bloc dont ils proviennent et (2) leur importance typographique. Ils forment finalement un effet visuel proche de celui du *tag cloud* (Figure 3). Etant donnée la structure arborescente d'une page Web, chacun des blocs peut récursivement contenir de nouveaux ensembles auxquels on peut appliquer ce procédé afin d'obtenir un arbre de *tag clouds* ; les feuilles de l'arbre pouvant être le niveau des paragraphes (correctement oralisés par une synthèse de la parole à partir de textes). L'idée sous-jacente est de proposer grâce au calcul de cette structure des stratégies de navigation intra-page innovantes.

Dans le cadre de la navigation non visuelle, nous souhaitons permettre une navigation intra-page oral-oral en produisant dynamiquement les versions audio des nœuds de la structure (que nous appelons par analogie *tag thunder*) ; ainsi que les moyens interactifs pour naviguer de l'une à l'autre. De la même manière que les mots d'un *tag cloud* sont distribués spatialement avec des effets visuels divers (taille, couleur…), les *tag thunders* distribuent les mots « spatiotemporellement » en jouant sur des effets sonores variés (volume, débit, hauteur, genre de la voix, fréquence de répétition, spatialisation). Les utilisateurs pourront interagir avec le système en prononçant un des mots-clés, appelé mot-clé de navigation (Voir Figure 3). L'automatisation de cette solution a nécessité la mise en place de l'architecture logicielle de la Figure 4, basée sur un plugin Firefox (non encore déployé), un serveur de synthèse de parole multicanal et un serveur pour le calcul du *tag thunder*.

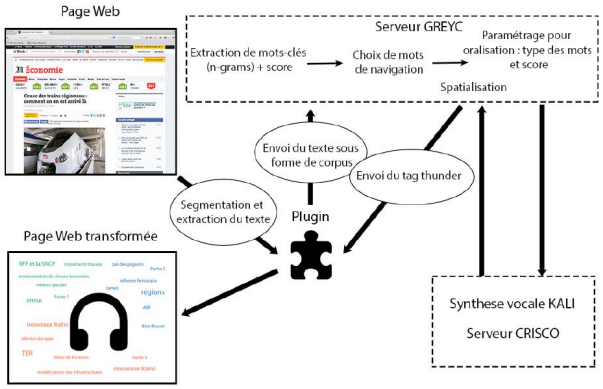

Figure 4. Schéma de l'architecture logicielle

Segmentation de la page en blocs : différentes approches existent. [11] propose de s'appuyer sur l'arbre DOM (Document Object Model) du HTML, [12] sur l'apparence de la page dans le navigateur (approche basée vision), [13] utilise des techniques de traitement d'image, [14] la structure sémantique ; [15] une résolution par graphe. Nous utilisons actuellement une méthode hybride que nous avons développée [16].

Résumé de la structure logico-thématique de la page : l'extraction d'expressions clés s'appuie sur des scores calculés conjointement sur le document et un corpus de référence. Cette mesure, appelée TF-IDF, permet d'étudier la pertinence des mots dans le contexte : la fréquence des termes dans le texte est pondérée par l'inverse de leur fréquence dans le corpus, faisant remonter les mots les plus spécifiques. Notre algorithme, inspiré de [17] s'appuie sur cette mesure combinée à l'indice de position du mot.

Interface de création de *tag thunder* : afin d'étudier de quelle manière les mots du *tag thunder* produits simultanément peuvent rester distinguables et informatifs, nous nous appuyons sur une interface que nous avons développée. Elle permet d'interroger le serveur de synthèse Kali du CRISCO [18] puis de faire varier les paramètres sonores. Nous pouvons ainsi créer des stimuli pour les expérimentations à venir. Il s'agira de parvenir à produire un effet « *cocktail party* » mis en évidence dans [19] : nous pouvons focaliser notre attention auditive sur un flux verbal dans une ambiance bruyante ; mais même si notre attention est fixée sur la parole d'un interlocuteur, nous restons dans une certaine mesure sensibles aux sons extérieurs.

Conclusions et perspectives

Outre les expérimentations évoquées dans le paragraphe précédent, beaucoup d'améliorations peuvent déjà être envisagées en parallèle. L'utilisation d'un corpus statique pose le problème du contexte : les mots apparus après 2006 sont inconnus et ont un IDF maximal même s'ils sont aujourd'hui communs. Le journal a un contexte langagier propre et malgré le grand nombre d'articles, cela peut rendre les résultats moins pertinents. L'IDF pourrait être calculé dynamiquement sur l'ensemble des pages qui composent le site de page cible (potentiellement combiné avec un corpus statique généraliste). Une autre amélioration en cours d'évaluation est l'utilisation de techniques d'écoute binaurales pour produire une spatialisation 3D dans un casque stéréo [20].

Références

[1] Dias G., Conde B. Accessing the web on handheld devices for visually impaired people. In K. Wegrzyn-Wolska and P. Szczepaniak, editors, Advances in Intelligent Web

Mastering, volume 43 of Advances in Soft Computing, 2007, pages 80-86.

[2] Ahmed F., Borodin Y., Soviak A., Islam M., Ramakrishnan I.V., Hedgpeth T. A*ccessible skimming: Faster screen reading of web pages*. In UIST 2012, 2012, pages 367-378.

[3] Borodin Y., Bigham J.P., Dausch G., Ramakrishnan I.V.. *More than meets the eye: A survey of screen-reader browsing strategies*. In International Cross Disciplinary Conference on Web Accessibility (W4A), 2010, pages 1–10.

[4] Bigham J.P., Cavender A.C., Brudvik J.T., Wobbrock J.O., Lander R.E. *Webinsitu: A comparative analysis of blind and sighted browsing behavior*. In 9th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS), 2007, pages 51–58.

[5] Ahmed F., Borodin Y., Puzis Y., Ramakrishnan L.V. *Why read if you can skim : towards enabling faster screen reading*. In W4A2012, 2012, Article No. 39.

[6] Maurel F., Vigouroux N., Raynal M., Oriola B. *Contribution of the Transmodality Concept to Improve Web Accessibility*. In Assistive Technology Research Series, Volume 12, 2003, Pages 186-193.

[7] Maurel F., Mojahid M., Vigouroux V., Virbel J.. *Documents numériques et transmodalité. Transposition automatique à l'oral des structures visuelles de texte.* Dans : Document numérique, Hermès, Vol. 9, N. 1, 2006, p. 25-42.

[8] Parmanto B., Ferrydiansyah R., Saptono A., Song L, Sugiantara I.W., Hackett S. *AcceSS: Accessibility through Simplification & Summarization*. In Proceedings of W4A, 2005, pages 18–25.

[9] Salampasis M., Kouroupetroglou C.. *Adaptive browsing shortcuts: Personalising the user interface of a specialised voice web browser for blind people*. In 23rd International Interconnect Technology Conference (IEEE), 2007, pages 818–825.

[10] Borodin Y., Ahmed F., Islam M.A., Puzis Y., Melnyk V., Feng S., Ramakrishnan L.V., Dausch G. *Hearsay: a new generation context-driven multi-modal assistive web browser*. In WWW'10, 2010, pages 1233–1236.

[11] Sanoja A., Gancarski, S. *Block-O-Matic: A web page segmentation framework*, Proceedings of ICMCS 2014, 2014, pp 595-600.

[12] Deng C., Shipeng Y., Ji-Rong W., Wei-Ying M. *VIPS : a Vision-based Page Segmentation Algorithm*. Microsoft Research, Technical Report MSR-TR-2003-79, 2003.

[13] Cao J., Mao B., Luo J. *A segmentation method for web page analysis using shrinking and dividing*. International Journal of Parallel, Emergent and Distributed Systems - Network and parallel computing,Volume 25 Issue 2, 2010, pages 93–104.

[14] Foucault N., Rosset S., Adda G. *Pré-segmentation de pages web et sélection de documents pertinents en Questions-Réponses*. TALN-RÉCITAL, 2013.

[15] Liu X., Lin H., Tian Y. *Segmenting Webpage with Gomory-Hu Tree Based Clustering*. Journal of Software, Vol 6, No 12, 2011, pages 2421–2425.

[16] Safi W., Maurel F., Routoure J.-M., Beust P., Dias G. A Hybryd Segmentation of Web Pages for Vibro-Tactile Access on Touch-Screen Devices, Proceedings of VL'14, associated to COLLING 2014, 2014, pp 95-102.

[17] Witten I.H., Paynter G.W., Frank E., Gutwin C., Nevill-Manning C.G. *KEA : Practical Automatic Keyphrase Extraction*. NRC/ERB-1057, 1999.

[18] Morel M., Lacheret-Dujour A. *Kali, synthèse vocale à partir du texte : de la conception à la mise en oeuvre*. Traitement Automatique des Langues 42, 2001, pages 193–221.

[19] Guerreiro J. *Using simultaneous audio sources to speed-up blind people's web scanning*. In 10th International Cross-Disciplinary Conference on Web Accessibility (W4A), pages 1–2, 2013.

[20] Nicol R., Gros L., Colomes C., Noisternig M., Warusfel O., Bahu H., Katz B. Simon L. *A Roadmap for Assessing the Quality of Experience of 3D Audio Binaural Rendering*. Proc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, 2014.