

Graphes typés pour l'exploration d'actualités

Rémi Bois*, Guillaume Gravier*, Emmanuel Morin**, Pascale Sébillot***

*CNRS, IRISA & INRIA Rennes, 263 Avenue Général Leclerc, 35042 Rennes
remi.bois@irisa.fr, guillaume.gravier@irisa.fr

**LS2N, Université de Nantes, 2 Chemin de la Houssinière, 44300 Nantes
emmanuel.morin@univ-nantes.fr

***INSA, IRISA & INRIA Rennes, 263 Avenue Général Leclerc, 35042 Rennes
pascale.sebillot@irisa.fr

1 Navigation dans des collections d'actualités

L'accès à l'information n'a jamais été aussi simple. Des dizaines d'articles sont créés chaque minute et rendus accessibles instantanément via les plateformes de médias en ligne. Néanmoins, confrontés à une telle masse d'informations disponibles, les professionnels peinent à rassembler efficacement l'ensemble des faits liés à un évènement. C'est le cas des journalistes, qui doivent mettre en contexte une information qu'ils s'appêtent à publier, ou des attachés de presse qui sont chargés de rassembler les informations disponibles à propos d'un sujet de société. Dans ces deux tâches, l'approche privilégiée consiste à utiliser un moteur de recherche afin de trouver les informations importantes liées à une thématique. Cette approche est cependant imparfaite, une recherche menant à des documents souvent très similaires, des pans entiers d'une information pouvant être introuvables sans l'utilisation du mot-clé pertinent.

D'autres outils permettent l'exploration de grandes collections d'actualités, tels que les frises chronologiques, l'organisation en fils d'actualités, ou le clustering d'informations. Ces outils sont néanmoins inadaptés à grande échelle, générant des frises chronologiques trop longues ou un nombre de clusters trop important. Dans le cadre du projet LIMAH, et en nous fondant sur des entretiens initiaux avec des professionnels afin de déterminer leurs besoins réels en termes de fonctionnalités (Gravier et al., 2016), nous proposons une représentation sous forme de graphe de l'ensemble des documents, avec des liens créés automatiquement entre des paires de documents d'actualité reliés sémantiquement¹. Ces liens sont typés via des heuristiques, permettant d'indiquer à l'utilisateur la nature de la relation entre deux documents.

2 Graphes typés : construction et évaluation par l'usage

L'interface réalisée consiste en un moteur de recherche et un outil de navigation. Après une première requête adressée à un moteur classique, l'utilisateur peut visualiser l'article de son

1. Les entretiens et tests utilisateurs ont été conduits avec l'aide de Éric Jamet et Maxime Robert, chercheurs au CRPCC de Rennes. L'interface a été réalisée avec Arnaud Touboulic, ingénieur.

Graphes typés pour l'exploration d'actualités

The screenshot displays the LIMAH interface for a news article. At the top, it shows the date 'May 29, 2015' and the article title 'FIFA : Joseph Blatter réélu pour un cinquième mandat'. The main content area (A) contains the full text of the article. Above the text is a network graph (C) with nodes representing related articles and their relationships. The graph shows a central node 'C' connected to nodes for 'Archives', 'Development', and 'D' (related articles). The interface also includes a sidebar with filters (E) for sources like 'Le Monde', 'Libération', and 'France Télévision', and a list of related articles (D) on the right. The top navigation bar includes search, favorites, and account options.

FIG. 1 – Interface LIMAH

choix et accéder à l'outil de navigation tel qu'il est présenté en figure 1. Celui-ci est composé d'une partie centrale dans laquelle l'article courant est visible dans son entièreté (A) ainsi que d'une zone affichant les articles liés (C), qui sont aussi fournis sous forme de liste (D). Des métadonnées sont également disponibles (B) telles que la date de publication et l'auteur de l'article actuel. Une liste de filtres (E) est proposée afin de permettre à l'utilisateur de masquer certaines sources. En double-cliquant sur l'un des nœuds du graphe de la zone C, l'utilisateur peut visionner un autre article et se déplacer au sein du graphe, mettant à jour chacune des parties de l'interface. Les nœuds sont organisés selon 7 types de relations : deux relations temporelles (antérieur - postérieur), deux relations de réaction à une information (réagit à - est source de réaction), et trois relations de similarité (résumé - développement - quasi duplicat).

L'intérêt de cette interface a été évalué lors de tests utilisateurs auprès d'étudiants en journalisme en fin d'études. Trois versions différentes ont été comparées : une version complète telle que présentée en figure 1, une version dans laquelle tous les liens sont de type temporels (les liens non temporels sont alors réaffectés à un type temporel), et une version ne proposant aucun article lié (zones C et D absentes). La tâche proposée aux utilisateurs a consisté à rassembler le maximum d'informations sur un sujet d'actualité imposé dans un temps limité. L'outil complet s'est révélé significativement plus efficace que les deux autres versions amputées testées, et a permis aux utilisateurs de retrouver davantage d'informations, y compris certaines n'apparaissant que dans quelques documents au sein de la collection. Les résultats obtenus ainsi que les entretiens réalisés avant et après les tests utilisateurs permettent de confirmer l'utilité de cette représentation à des fins d'exploration de collections d'actualités.

Références

Gravier, G., M. Ragot, L. Amsaleg, R. Bois, G. Jadi, E. Jamet, L. Monceaux, et P. Sébillot (2016). Shaping-up multimedia analytics : Needs and expectations of media professionals. In *Proceedings of the 22nd MMM Conference, Perspectives on Multimedia Analytics*.