

Une interface pour la validation et l'évaluation de chronologies thématiques

Xavier Tannier^{1,2} Véronique Moriceau^{1,2} Erwan Le Flem^{1,3}

(1) LIMSI-CNRS

(2) Université Paris-Sud, 91403 Orsay, France

(3) IUT de Vannes

prenom.nom@limsi.fr

RÉSUMÉ

Cet article décrit une interface graphique de visualisation de chronologies événementielles construites automatiquement à partir de requêtes thématiques en utilisant un corpus de dépêches fourni par l'Agence France Presse (AFP). Cette interface permet également la validation des chronologies par des journalistes qui peuvent ainsi les éditer et les modifier.

ABSTRACT

An Interface for Validating and Evaluating Thematic Timelines

This demo paper presents a graphical interface for the visualization and evaluation of event timelines built automatically from a search query on a newswire article corpus provided by the Agence France Presse (AFP). This interface also enables journalists to validate chronologies by editing and modifying them.

MOTS-CLÉS : chronologie événementielle, évaluation, validation.

KEYWORDS: event timeline, evaluation, validation.

1 Construction des chronologies événementielles

Actuellement, les journalistes de l'Agence France Presse (AFP) construisent manuellement des chronologies événementielles textuelles dans le but de contextualiser des événements médiatiques. Elles sont sous forme d'une liste de dates (généralement entre 10 et 20) associées à un texte décrivant l'événement ayant eu lieu à cette date.

Nous avons développé un système capable de construire ce genre de chronologies événementielles thématiques à partir d'une requête utilisateur en utilisant un corpus de dépêches en français et en anglais fourni par l'AFP¹. Nous considérons que les événements importants, ceux que nous souhaitons retrouver dans les chronologies, ont lieu à des dates importantes (du point de vue du thème imposé par la requête de l'utilisateur). Pour extraire les dates qui méritent de figurer dans une chronologie événementielle, les expressions temporelles dans les textes sont dans un premier temps reconnues et normalisées. Nous utilisons ensuite une approche par apprentissage pour extraire les dates saillantes pour un thème donné (Kessler *et al.*, 2012). En sortie, notre

1. Ce travail a été partiellement financé par l'ANR dans le cadre du projet Chronolines (ANR-10-CORD-010). Nous remercions l'AFP pour la mise à disposition du corpus et son concours pour la définition de l'interface.

système fournit une liste de dates (entre 20 et 300) classées de la plus à la moins importante par rapport au thème de la requête. Chaque date est accompagnée d'une description d'une phrase ainsi que d'un ensemble de phrases pertinentes pour l'événement décrit.

2 Interface de validation des chronologies

L'interface graphique comporte 3 vues (voir figure 1). La partie ① présente la chronologie construite automatiquement en fonction de la requête de l'utilisateur. L'importance de l'événement est indiquée par un code couleur. La partie ② permet à l'utilisateur de valider manuellement la chronologie. Il peut sélectionner dans ①, par glisser-déplacer, les événements qui lui semblent pertinents. Ceux-ci sont modifiables si besoin (date et description textuelle de l'événement). L'utilisateur peut également créer de nouveaux événements qui ne seraient pas présents dans les propositions de la chronologie automatique. Enfin, la partie ③ donne une visualisation de la chronologie validée sous forme de chronologie navigable. Elle est implémentée avec l'outil web SIMILE TimeLine, une API JavaScript permettant de générer des chronologies visuelles sur une page HTML. Cette interface permet non seulement de faciliter le travail des journalistes dans la construction de chronologies mais aussi d'évaluer la qualité des chronologies produites automatiquement en les comparant à celles produites par les journalistes.

The image shows a web interface for timeline validation, divided into three main sections:

- Top Section (3):** A timeline visualization titled "Visualisation des chronologies". It shows a horizontal timeline from December 19 to March 20, 2011. Key events are marked with colored dots and labels: "President Hosni Mubarak" (blue), "January 29, 2011:" (blue), "Egyptian Vice President" (blue), "Egypt's embattled" (blue), and "Tens of thousands of" (blue).
- Middle Section (2):** A validation interface titled "Validation des chronologies". It features a search bar with the query "mubarak egypt" and a date range from "2011/01/01" to "2011/06/30". Below the search bar, there are buttons for "Open find bar", "New event", "Delete selection", and "Change status". A list of events is displayed, including "2011/01/25 - January 25, 2011 : First major protests against Mubarak" and "2011/02/03 Right now I care about my country, I care about Egypt, he said on Thursday".
- Bottom Section (1):** A list of events titled "Validation des chronologies". It shows a list of events with their dates and descriptions, such as "2011/01/25 - January 25, 2011 : First major protests against Mubarak" and "2011/02/03 Right now I care about my country, I care about Egypt, he said on Thursday".

FIGURE 1 – Interface de visualisation et de validation des chronologies.

Références

KESSLER, R., TANNIER, X., HAGÈGE, C., MORICEAU, V. et BITTAR, A. (2012). Finding Salient Dates for Building Thematic Timelines. *In 50th Annual Meeting of the ACL*, République de Corée.