

# De quoi parle ce Tweet ? Résumer Wikipédia pour contextualiser des microblogs

Romain Deveaud<sup>1</sup>, Florian Boudin<sup>2</sup>

<sup>1</sup> University of Glasgow, UK

`romain.deveaud@glasgow.ac.uk`

<sup>2</sup> LINA - UMR CNRS 6241, Université de Nantes, France,

`florian.boudin@univ-nantes.fr`

**Résumé:** Les réseaux sociaux sont au centre des communications sur internet et un grand parti des échanges communautaires se fait à travers eux. Parmi eux, l'apparition de Twitter a donné lieu à la création d'un nouveau type de partage d'informations où les messages sont limités à 140 caractères. Les utilisateurs de ce réseau s'expriment donc succinctement, souvent en temps réel à partir d'un smartphone, et la teneur des messages peut parfois être difficile à comprendre sans contexte. Nous proposons dans cet article une méthode permettant de contextualiser automatiquement des Tweets en utilisant des informations provenant directement de l'encyclopédie en ligne Wikipédia, avec comme but final de répondre à la question : « De quoi parle ce Tweet ? »\_ Nous traitons ce problème comme une approche de résumé automatique où le texte à résumer est composé d'articles Wikipédia liés aux différentes informations exprimées dans un Tweet. Nous explorons l'influence de différentes méthodes de recherche d'articles liés aux Tweets, ainsi que de plusieurs caractéristiques utiles pour la sélection des phrases formant le contexte. Nous évaluons notre approche en utilisant la collection de la tâche Tweet Contextualization d'INEX 2012 et donnons un aperçu sur ce qui caractérise une phrase importante pour déterminer le contexte d'un Tweet.

**Mots clés:** Contexte informatif, résumé par extraction, Twitter, Wikipédia