

Texte et Représentation en Recherche d'Information

par Benjamin Piwowarski, Université Pierre et Marie Curie

Résumé

Les performances des modèles de Recherche d'Information (RI) sont fortement liées à leur capacité à représenter le texte. Ces dernières années, portés par les succès en reconnaissance d'image et de parole, de nombreux modèles basés sur les réseaux de neurones et les représentations distribuées des mots et phrases sont apparus. Quelques modèles ont essayé d'aborder le problème de la RI, dont une des difficultés principale est de demander une représentation à la fois précise (pour les entités) et diffuses (pour les concepts). Dans cette présentation, j'introduirai d'abord les problématiques de représentation en RI, les développements récents dans le domaine des réseaux de neurones, présenterai mes travaux en cours portant sur la prise en compte du contexte et du temps, et discuterai de quelques pistes de travail dans ce domaine.

Bio

Benjamin Piwowarski est chargé de recherche 1ère classe du CNRS au LIP6 (UPMC). Son domaine de recherche est celui de l'accès à l'information et de l'apprentissage statistique appliqué, en particulier l'apprentissage de représentations. Il a travaillé sur la recherche d'information XML (thèse 1999-2003), sur les bases de données XML (post-doctorat 2004-06 à l'université du Chili), sur l'analyse des interactions entre un utilisateur et un moteur de recherche (Yahoo! 2006-08) et sur l'application du formalisme mathématique de la physique quantique aux modèles d'accès à l'information (2008-11). Ses intérêts actuels portent sur (l'apprentissage de) représentation des données textuelles de graphes. Il a été fortement impliqué dans l'initiative INEX (2004-07), et a organisé un workshop (IRGM'07), deux tutoriels (ECIR 2012 et ICTIR 2013), et a participé l'organisation de CORIA 2015.