METSEM#12

SÉMINAIRE DE MÉTHODOLOGIE

Statistique textuelle

un état des lieux

Vendredi 23 mars - 10h30 à 12h30
Sciences Po, Salle Percheron
98 rue de l'Université, 75007 Paris

Bénédicte Garnier, INED

Cette présentation abordera, sous forme de retour d'expérience, les avancées de la statistique textuelle en termes de données, méthodes et outils.

L'essor du Big data et des données collectées sur le web a démultiplié les corpus : des données non traitées hier sont maintenant sujettes à analyses et questions. Les corpus exploitables sont de nature très variée : mots associés, titres d'articles, flux RSS, etc. Les bases en ligne et les techniques d'aspiration ont favorisé l'accès à d'autres corpus. De nouveaux enjeux apparaissent.

Aujourd'hui, on se doit de tout analyser. Cette « fouille » de textes ne dispense pas d'une connaissance de la thématique abordée et des données mises en jeu. La statistique textuelle permet, au moyen de méthodes sans a priori, de faire émerger les univers lexicaux d'un corpus ou de les rattacher à des données de contexte. Elle est aussi mobilisée en complément de méthodes qualitatives.

Tout comme l'analyse des réponses à une question ouverte, la visualisation de mots en nuages ou en graphes est une pratique qui devient courante. La statistique textuelle est favorisée par l'accès à des logciels gratuits ou libres.

Cet ensemble de nouveautés a permis d'élargir le champ des utilisateurs et le développement d'applications.

Inscription obligatoire
sur metsem.hypotheses.org

SciencesPo

cnrs