



Web search

Hot Topics in Web Search





New Aspects of Web Search

- **Web 2.0**: interaction-rich interfaces, social networks, mash-ups, user-provided content
- **Deep Web**: data accessible only through HTML forms
- **Information Extraction**: search at the level of objects inside pages
- **Semantic Web**: machine-readable content, complex search queries
- **Distributed computing**: huge volume of data in search engines' indexes; techniques for processing it, computing statistics, etc.

19 March 2010





Web 2.0

Definition

Web 2.0: buzzword about:

- **rich dynamic interfaces**, especially with the help of **AJAX** (Asynchronous JavaScript and XML) technologies: GMail, Google Suggest
- **user-editable** content, **collaborative** work and **social networks**: blogs, Wikipedia, MySpace, Facebook
- **aggregation** of content from multiple sources and **personalization**: Netvibes, Yahoo! Pipes





Web 2.0 Impact on Web Search

- How to crawl AJAX Web sites?
- Previous lecture about social networks and search in social networks
- How to exploit user feedback in search engines: **relevance feedback**, cf. slides by C. Manning and P. Raghavan, Stanford:
<http://www.stanford.edu/class/cs276/handouts/lecture9-queryexpansion-1up.pdf>





Deep Web

cf. slides on the course website:

<http://pierre.senellart.com/enseignement/2009-2010/inf396/5-hot-topics/deep-web.pdf>





Information Extraction

cf. course by William Cohen, CMU:

<http://www.cs.cmu.edu/~wcohen/10-707/01-16-introduction.ppt>





cf. online tutorial by Ivan Herman:

<http://www.w3.org/2006/Talks/0428-Beijing-IH/>





Distributed Computing

cf. slides on the course website:

<http://pierre.senellart.com/enseignement/2009-2010/inf396/5-hot-topics/sldistr.pdf>





Licence de droits d'usage



Contexte public } avec modifications

Par le téléchargement ou la consultation de ce document, l'utilisateur accepte la licence d'utilisation qui y est attachée, telle que détaillée dans les dispositions suivantes, et s'engage à la respecter intégralement.

La licence confère à l'utilisateur un droit d'usage sur le document consulté ou téléchargé, totalement ou en partie, dans les conditions définies ci-après et à l'exclusion expresse de toute utilisation commerciale.

Le droit d'usage défini par la licence autorise un usage à destination de tout public qui comprend :

- le droit de reproduire tout ou partie du document sur support informatique ou papier,
- le droit de diffuser tout ou partie du document au public sur support papier ou informatique, y compris par la mise à la disposition du public sur un réseau numérique,
- le droit de modifier la forme ou la présentation du document,
- le droit d'intégrer tout ou partie du document dans un document composite et de le diffuser dans ce nouveau document, à condition que :
 - L'auteur soit informé.

Les mentions relatives à la source du document et/ou à son auteur doivent être conservées dans leur intégralité.

Le droit d'usage défini par la licence est personnel et non exclusif.

Tout autre usage que ceux prévus par la licence est soumis à autorisation préalable et expresse de l'auteur : sitopedago@telecom-paristech.fr

19 March 2010

