

Web Information Retrieval: Programma preliminare (A.A. 05/06)

6 crediti

Docente

Stefano Mizzaro

Finalità del corso

Il corso mira a introdurre le tematiche principali dell'information retrieval (IR), una disciplina importante storicamente e che ha ricevuto un forte impulso (sia scientifico sia economico) in seguito all'avvento del Web. Infatti, dopo la posta elettronica, l'uso di un motore di ricerca è oggi la seconda attività degli utenti del Web; svariati utenti del Web usano un motore di ricerca come Google come modalità principale di accesso alla rete; e molte delle ricerche più avanzate vengono oggi effettuate nei laboratori privati (Google, Yahoo!, Microsoft). Oltre alle tematiche associate dell'IR classico, vengono presentate quelle, più recenti, specifiche del mondo Web. Vengono discussi sia argomenti di base sia le correnti linee di ricerca e le tendenze future.

Tematiche

Il corso è diviso in tre parti:

1. Information Retrieval classico: modelli concettuali della ricerca d'informazioni, modelli concettuali e formali dei sistemi d'IR, progetto e implementazione di un sistema d'IR, strutture dati e algoritmi di indicizzazione e ricerca, interfacce utente per l'IR, la problematica della valutazione dei sistemi d'IR.
2. Web Information Retrieval: modelli del Web (forma e dimensioni del grafo del Web, cenni all'analisi delle reti sociali e alle reti piccolo mondo), analisi dei link e della connettività, iperlink e crawling, IR multimediale, meta-motori, progetto e implementazione di un motore di ricerca per il Web, biblioteche digitali.
3. Tematiche avanzate: IR parallelo e distribuito, ricerca d'informazioni tramite dispositivi mobili e su reti peer-to-peer, problemi fondazionali del settore (ad es., i concetti di informazione e pertinenza), sistemi per il filtraggio d'informazioni, clustering, banche dati specialistiche (ad es., mediche), ecc. Queste tematiche avanzate verranno trattate in forma di attività seminariale, che potrà essere svolta, opzionalmente, dagli studenti.

Modalità d'esame

Orale. Sono previste anche attività, opzionali e a scelta dello studente ma caldamente consigliate, che portano a un esame semplificato e a una votazione superiore: seminari svolti durante il corso o semplici progetti concordati con il docente. Al corso è possibile affiancare il laboratorio: in tal caso l'attività progettuale è più impegnativa.

Testi consigliati

- R. Baeza-Yates e B. Ribeiro-Neto, *Modern Information Retrieval*, Addison Wesley, 1999. ISBN: 020139829X.
- R. Belew, *Finding Out About - A Cognitive Perspective on Search Engine Technology and the WWW*, Cambridge University Press, 2000, ISBN: 0-521-63028-2.
- R. Korfhage, *Information Storage and Retrieval*, Wiley, 1997, ISBN: 0471143383.

Molti degli argomenti trattati non sono disponibili su testi, ma solo su articoli scientifici. A lezione verranno comunicati eventuali altri testi e verrà segnalato o fornito altro materiale (tutti i lucidi delle lezioni, pagine Web, dispense, altri testi di riferimento, articoli scientifici, ecc.).

Web Information Retrieval: Preliminary Program (2005/06)

6 credits

Lecturer

Stefano Mizzaro

Aims

To present the main topics in Information Retrieval (IR), a subject that has a high historical importance and has received an even increased attention (from both the scientific and economic viewpoints) after the coming of the Web. Indeed, after e-mail, using a search engine is today the second activity of Web users; several Web users perceive a search engine (typically, Google) as the main access to the Web; and several advanced research topics are studied in commercial laboratories (Google, Yahoo!, Microsoft). Besides classical and well established IR topics, the course emphasizes more recent Web specific issues. Both basic topics and advanced research trends are discussed.

Topics

The course is divided into three parts:

1. Classic Information Retrieval: conceptual models of information seeking, conceptual and formal models of IR systems, design and implementation of IR systems, data structures and algorithms for indexing and search, user interfaces for IR, evaluation of IR systems.
2. Web Information Retrieval: Web modeling (shape and size of the Web graph, introduction to social network analysis and small world networks), link and connectivity analysis, hyperlink and crawling, multimedia IR, meta-search, design and implementation for Web search engines, digital libraries.
3. Advanced issues: parallel and distributed IR, information seeking with mobile devices and on peer-to-peer networks, foundational issues (e.g., concepts of information and relevance), information filtering and routing, clustering, specialized (e.g., medical) databases, etc. These advanced issues are presented by means of seminars by invited experts or, optionally, by students.

Exams

Oral examination plus optional (but warmly suggested) term projects or seminars, that lead to a simpler oral exam and to higher marks. A laboratory can be attached to the course, in which case the project and/or seminar is more demanding.

Textbooks

- R. Baeza-Yates e B. Ribeiro-Neto, *Modern Information Retrieval*, Addison Wesley, 1999. ISBN: 020139829X.
- R. Belew, *Finding Out About - A Cognitive Perspective on Search Engine Technology and the WWW*, Cambridge University Press, 2000, ISBN: 0-521-63028-2.
- R. Korfhage, *Information Storage and Retrieval*, Wiley, 1997, ISBN: 0471143383.

Further textbooks may be communicated at course start. Other teaching material (slides covering all the lectures, Web pages, lecture notes, scientific articles, reference texts) will be communicated or provided during the lectures.