

A definition of IR [Manning et.al. 2007]

Information retrieval (IR) is finding material (usually documents) of unstructured nature (usually text) that satisfy an information need from within large collections (usually on local computer servers or on the internet).

Dip. di Matematica Pura ed Applicata F. Aiolli - Sistemi Informativi 2007/2008

- 1. Finding material (documents) in *large* collections
- Unstructured nature as opposed to structured data: data do not have a fixed semantic/structure
- 3. Satisfy an *information need*

3



Dip. di Matematica Pura ed Applicata F. Aiolli - Sistemi Informativi 2007/2008

	A desire (possibly specified in an imprecise way) of information useful to the solution of the problem, or resources useful to a given goal;
	Useful (Relevant), according to the subjective opinion of the user.
Dip. Pura	di Matematica F. Aiolli - Sistemi Informativi 7 ed Applicata 2007/2008
Ti	pical tasks covered in IR
Ti	pical tasks covered in IR Search ('ad hoc' retrieval) Static document collection,
Ti	pical tasks covered in IR Search ('ad hoc' retrieval) Static document collection, Dynamic queries Filtering Static query, Dynamic document feeds
	pical tasks covered in IR Search ('ad hoc' retrieval) Static document collection, Dynamic queries Filtering Static query, Dynamic document feeds Categorization Clustering
	pical tasks covered in IR Search ('ad hoc' retrieval) • Static document collection, • Dynamic queries Filtering • Static query, • Dynamic document feeds Categorization Clustering Collaborative Filtering or Recommendation Browsing Summarization



	Data Retrieval, as in DBs
	Information need cannot directly be expressed as a simple query and documents have not a precise semantic. A translation of them into logical representations is needed
	 In IR the set of objects to be retrieved are not clearly determined -> Slightly different retrieved sets should not be necessarily considered as a 'fatal' error of the system User satisfaction is the issue of IR
	 Knowledge Retrieval, as in AI In AI a fact α is inferred from a knowledge base Γ of facts expressed in a certain formalism
	 Question Answering In QA a query an answer is returned generated from a semantic analysis of documents. Huge amount of domain knowledge needed
	 Information Browsing, as in Hypermedia Relevant documents are retrieved by an active intervention of the user and not by a search routine
	The goal of a browsing task is less clear in the mind of the user
Dip	o. di Matematica F. Aiolli - Sistemi Informativi 11
E	Evolution of IR In the past, IR systems were used only by expert librarians as reference retrieval systems in batch modality.
E	Evolution of IR In the past, IR systems were used only by expert librarians as reference retrieval systems in batch modality. Many libraries still use categorization hierarchies to classify their volumes The advent of novel computers and the Web have brought to

- interaction with the user
- methods to deal with distributed document collections (e.g. WWW)

Dip. di Matematica Pura ed Applicata

	An IR model (can be defined by M=[D,Q,R] wher	2
	D is a repr	esentation for the documents in the	collection
	Q is a repr (queries)	resentation for the user informatior	needs
	R(d _i ,q _j) is a a real num representa	a ranking function which associates ber with a query $\mathbf{q_j} \in \mathbf{Q}$ and a docum ation $\mathbf{d_i} \in \mathbf{D}.$	ent
	N.B. It defi a given	nes an ordering among the documents w query q _j .	ith regard to
Dip. d Pura e	li Matematica ed Applicata	F. Aiolli - Sistemi Informativi 2007/2008	13
	Unlike query sati between docume but is subjective	sfaction in DBs, the relationship of <i>releva</i> nts D and information needs Q is not form , i.e. determined by the user. Therefore,	<i>nce</i> R ally defined,
	Unlike query sati between docume but is subjective unlike in DBs, a degree of en	isfaction in DBs, the relationship of <i>releva</i> nts D and information needs Q is not form , i.e. determined by the user. Therefore, effectiveness is an issue ffectiveness (user satisfaction) can be defined	<i>nce</i> R ally defined,
	Unlike query sati between docume but is subjective unlike in DBs, a degree of en In an IR system relevance as	sfaction in DBs, the relationship of <i>releva</i> nts D and information needs Q is not form , i.e. determined by the user. Therefore, effectiveness is an issue ffectiveness (user satisfaction) can be defined or model, it is necessary to choose whethe	<i>nce</i> R ally defined, ir to treat
	Unlike query sati between docume but is subjective unlike in DBs, a degree of et In an IR system relevance as 1. Boolean-valued 2. Finite-valued 3. Infinite-value	isfaction in DBs, the relationship of <i>releva</i> nts D and information needs Q is not form , i.e. determined by the user. Therefore, effectiveness is an issue ffectiveness (user satisfaction) can be defined or model, it is necessary to choose whethe d R: $D \rightarrow Q \in \{0,1\}$ R: $D \rightarrow Q \in \{1,,N\}$ d R: $D \rightarrow Q \in \mathbb{R}$	<i>nce</i> R ally defined, er to treat

