

02-28-2024

INFORMATION RETRIEVAL

Dichiarazione universale dei diritti umani: "diritto di cercare, ricevere e dare informazioni tramite ogni media e indipendentemente da confini"

Sia ricerche precise che ricerche vaghe -> per ricerche precise, possibile trovare risposta esatta (fattuale), anche se dati possono renderlo difficile

Strutture dati definiscono: dominio, operazioni possibili, codifica -> per questo, database funzionano meglio per IR piuttosto che stringhe di testo -> es: per classi in Java definiamo queste tre cose (metodi, attributi, etc.)

Gerald Salton: IR riguarda struttura, analisi, organizzazione, salvataggio, ricerca e ottenimento di informazioni -> bisogna trovare struttura dati da documenti per permettere ricerca e ottenimento (database: definita prima)

Focus primario dall'inizio: testo e documenti -> possono avere organizzazione

Tipi di task:

- AD-HOC SEARCH: trovare documenti su certo topic (homework)
- KNOWN ITEM SEARCH: certo elemento (domande fattuali)
- ANSWER SEEKING: certa domanda
- DIRECTED EXPLORATION: e.g. lista di elementi
- DECISION MAKING
- EXPERT SEARCH

Confronto tra IR e database: collezione di dati correlati -> dati: fatti noti che possono essere registrati e hanno significato implicito (pezzi atomici di informazioni che insieme fanno dati correlati; da soli non molto utili) -> DBMS: sistema computerizzato per creare e gestire database

Semantiche di ogni pezzo di informazione ben definiti -> attributi ben definiti

03-01-2024

Database: dati strutturati con significato noto

DATA INFORMATION KNOWLEDGE AND WISDOM pyramid

(<https://www.ontotext.com/knowledgehub/fundamentals/dikw-pyramid/>): dati sono fatti atomici, poi li combiniamo in database, così otteniamo informazione, da quella otteniamo conoscenza
Conoscenza costruita da computer: e.g. data mining