



Università degli studi di Siena

Corso di laurea in Scienze della Comunicazione

La valutazione dei sistemi di QUESTION – ANSWERING

Tesi di laurea in Linguistica Computazionale

Candidato: Chiara Brinchi Giusti

Relatore: Amedeo Cappelli

OBIETTIVO DEL LAVORO

Analizzare la struttura e il
funzionamento dei sistemi di
QUESTION-ANSWERING (Q/A)...

..e le tecniche usate in **TREC**
(Text Retrieval Conference) per
valutare le loro prestazioni.

PERCHE' QUESTA SCELTA ?

- Crescita del numero di persone che quotidianamente ha la necessità di ricercare informazioni contenute in grandi archivi di dati (Internet e moderni database).
- Problema della gestione e del recupero (tramite interrogazione in LN) di informazioni rilevanti.
- Diffusione di ricerche volte a sviluppare sistemi di recupero dinamici, veloci ed estremamente precisi.

COSA SONO I SISTEMI DI Q/A?

"Software di recupero automatico delle informazioni che rispondono a domande formulate da un utente in Linguaggio Naturale"

- Forniscono risposte estremamente precise.
- Lavorano su grandi quantità di dati (Open-domain Vs Limited-domain Vs Closed-domain).
- Integrano le classiche tecniche di IR con metodi di estrazione della risposta che si basano sulla disambiguazione semantica della domanda (Query).
- FALCON: è il sistema che ha ottenuto i risultati migliori nella sezione del Q/A di Trec-9 (Southern Methodist University di Dallas, USA, 1999).

STRUTTURA IN TRE MODULI:

• **QUESTION PROCESSING:** parsing sintattico dei termini della domanda e loro successiva trasformazione in forma semantica grazie all'uso di database lessicali come ad esempio WordNet, utilizzato da Falcon.

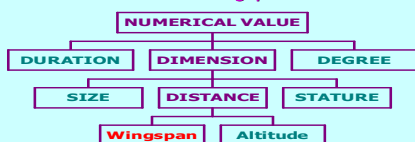
• **PARAGRAPH PROCESSING:** ricerca della risposta all'interno dei documenti attraverso un meccanismo di recupero strutturato in tre cicli e basato su un processo di espansione dei termini più significativi della Query.

• **ANSWER PROCESSING:** estrazione della risposta, seguita dalla fase della valutazione dei risultati (*P.Evaluation*).

ANALISI DELLA DOMANDA:

- **Parsing sintattico:** ogni elemento della Query viene associato alla categoria lessicale che gli appartiene.
- **Forma Semantica:** ogni parola viene associata ad un'etichetta di dominio (Distanza, Tempo, Storia) che permette di creare un diagramma basato sulla similarità semantica tra le parole.
- **WordNet:** database lessicale organizzato su base concettuale e non alfabetica.

Es: "What is the wingspan of a condor?"



TRATTAMENTO DEI DOCUMENTI ED ESTRAZIONE DELLA RISPOSTA:

- **Dispositivo contenuto all'interno di un meccanismo di ricerca strutturato in tre cicli:** espansione della Query → aggiunta di nuovi termini alla domanda (si utilizza solo se il sistema non è ancora in grado di recuperare la risposta attesa).

Es: "Chi fu l'inventore della lampadina?"

- 1° ciclo: derivazione MORFOLOGICA → estrazione (insuccesso)
↓
2° ciclo: derivazione LESSICALE → estrazione (insuccesso)
↓
3° ciclo: derivazione SEMANTICA → estrazione (nuovo ciclo)

I sistemi di Q/A fanno parte di un settore di ricerca estremamente competitivo, come dimostrano le innumerevoli conferenze che si svolgono ogni anno negli USA: una delle principali è TREC.

COS'E' TREC?

- TEXT RETRIEVAL CONFERENCE -

- E' una conferenza annuale che incoraggia la ricerca nel campo dei sistemi di estrazione di informazione da testo attraverso la premiazione dei software migliori.
- Favorisce il dialogo tra università ed aziende private (IBM, Xerox, AT&T Research...) attraverso un forum dove vengono scambiati progetti di ricerca.

PERFORMANCE EVALUATION

- Perché la fase della valutazione è importante?
- Valutare le prestazioni di un sistema di Q/A significa verificare se il software fornisce prestazioni adeguate rispetto agli obiettivi per cui è stato progettato. Per costruire un sistema di valutazione adatto al tipo di analisi che si intende effettuare bisogna:

DECIDERE QUALI ATTRIBUTI DEL SISTEMA VALUTARE.
(velocità elab. - numero di risp. corrette - % errori commessi)
STABILIRE CHE TIPO DI MISURAZIONI EFFETTUARE
TENENDO CONTO DEGLI ATTRIBUTI SCELTI.
EFFETTUARE LE MISURAZIONI E COMPARARE I
RISULTATI OTTENUTI DA DIVERSI SISTEMI.

DOCUMENTI E TIPO DI MISURAZIONI

- 500 domande (ad esclusione di domande SI/NO, procedurali e troppo attuali).
- Acquaint Corpus (3 gigabytes per 1.033.000 doc.)
- In risposta: lista ordinata di 5 frammenti testuali con una lunghezza di 50 o 250 caratteri.
- Esaminatore umano (Assessor).
- **MEAN RECIPROCAL RANK (MRR)**: l'esaminatore stabilisce qual è il frammento più corretto tra i 5, e gli assegna un punteggio pari al reciproco del numero di posizione in cui è stato riportato.
- **PRECISION**: rapporto tra il numero di risp. corrette estratte e il numero totale di risposte estratte.
- **RECALL**: rapporto tra il numero di risp. corrette estratte e il numero di risp. corrette da estrarre.

CONCLUSIONI

- Da anni ampi settori della ricerca universitaria, ma soprattutto le aziende private, investono nel settore di ricerca del Q/A cospicue energie. Questo interesse è legato principalmente ad una motivazione di carattere commerciale: infatti la possibilità di recuperare informazione selettiva a partire da archivi di dati non strutturati, elimina del tutto i costi di una preventiva strutturazione del database.
- Obiettivo futuro: riuscire ad eliminare l'intervento umano nella fase di estrazione/valutazione delle risposte; questo implica che i sistemi dovrebbero essere capaci di stabilire da soli il grado di rilevanza delle risposte restituite.