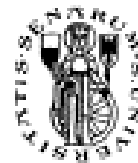


Question Answering e semantica

*Tecnologie emergenti
per le imprese*

Associazione Websemantico e Personae organizzano:



Cos'è QuestIT?



QuestIT nasce nell'ambito di attività di ricerca in Intelligenza Artificiale presso il Dipartimento di Ingegneria dell'Informazione, Università di Siena.

Il focus di questa attività è legato soprattutto allo sviluppo di tecnologie per il **Question Answering**. Grazie alla collaborazione con Expert System S.p.A., società leader nella tecnologia semantica, QuestIT si propone di diventare una realtà di riferimento nel mercato italiano ed internazionale.

L'innovatività dell'impresa è garantita dallo spirito d'iniziativa e dal know-how del team, che ha dimostrato nel recente passato di poter dar vita ad un progetto di ricerca, **WebCrow**, di ampio richiamo internazionale. Tale progetto (<http://webcrow.dii.unisi.it>) ha vinto per due anni (2006 e 2007) il Google Project Award.

Cos'è il Question Answering?

- “The goal [...] is [...] to retrieve answers rather than documents in response to a question”*.
- La differenza sostanziale rispetto all'Information Retrieval o Document Retrieval dei motori di ricerca è determinata dalla tipologia di informazione che l'utente cerca:
 - Risposta specifica ad una domanda che può essere anche molto complessa: “Essendo un lavoratore dipendente, entro quando devo decidere la destinazione del mio tfr maturato? ”
 - Informazioni generiche su un'argomento
- Il Question Answering (Q&A) consiste di una fase di Information Retrieval e di una detta di Answer Extraction dai documenti, che diventa realmente rilevante nei casi in cui l'utente cerca una risposta puntuale alla sua domanda.

* Ellen M. Voorhees, National Institute of Standards and Technology, “*Overview of the TREC 2003 Question Answering Track*”

Cos'è il Question Answering? [2]

- Open Domain
 - domande su **qualsiasi argomento**.
 - La base di conoscenza da cui il sistema estrae le informazioni deve avere una natura enciclopedica, pensiamo al Web, Wikipedia etc.
 - Spesso i sistema Open Domain si basano sul principio di ridondanza nei risultati e estraggono le risposte mediante tecniche statistiche.
- Specific Domain
 - Domande su un **argomento specifico** (es: La riforma della previdenza complementare)
 - La base di conoscenza ha dimensioni molto limitate (da 1 documento a poche centinaia), ma ha un topic omogeneo.
 - Mancanza di ridondanza.
 - E' possibile eseguire un'analisi più approfondita per modellare le caratteristiche linguistiche-semantiche del corpus, riuscendo a coprire la "distanza linguistica" tra il contenuto del corpus e le domande degli utenti

Q&A per domini ristretti



- QuestIT sviluppa sistemi di Question Answering (Q&A) per domini ristretti derivante dalla collaborazione con Expert System.
- L'obiettivo è di **rispondere in maniera sintetica e precisa** (evitando di presentare interi documenti) a domande formulate in linguaggio naturale.
- Sinergia tra **tecnologia linguistica**, tecniche statistiche ed Information Retrieval.
- Il valore aggiunto delle tecnologie di Q&A, rispetto a search engine classici, è **più evidente su corpus orientati ad utenti non esperti**, che usano un linguaggio comune per riferirsi a concetti specifici della base di conoscenza.

Q&A per domini ristretti [2]

- Il sistema effettua automaticamente una **parcellizzazione della documentazione**.
- Vengono **individuati porzioni di testo** con la relativa informazione di contesto: documento da cui è stato estratto, capitolo o sezione di appartenenza.

BlackBerry basics

About typing input methods

You can type on your BlackBerry® device using the SureType™ input method or the multi-tap input method.

SureType technology combines a traditional phone number key layout with a familiar computer-style letter layout that is designed to provide a comfortable typing and dialing experience.

When you use SureType technology, you press the letter key for each letter in a word once. For example, to type **run**, you would press the **[r]** key once, the **[u]** key once, and the **[n]** key once. As you type, a list of possible letter combinations and words appears on the screen. SureType technology selects letter combinations or words from the list based on context. If you try to type a word or letter combination (for example, a web address or an abbreviation) that SureType technology does not recognize, you can use the list that appears on the screen to build the word letter by letter. SureType technology is designed to "learn" new words or letter combinations that you type and add them to your custom word list.

When you use multi-tap, you press a letter key once to type the first letter on the key and twice to type the second letter. For example, to type **run**, you would press the **[r]** key twice, the **[u]** key once, and the **[n]** key twice.

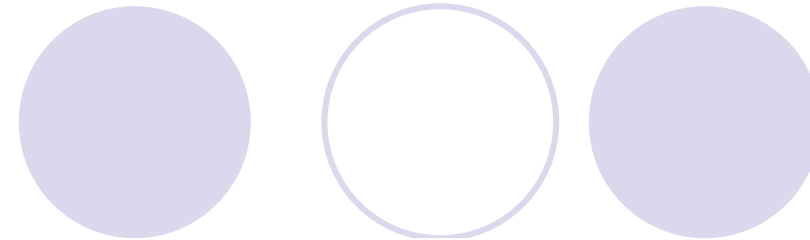
SureType technology is the default typing input method for most fields. Multi-tap is the default for phone number and password fields.

Type text using SureType technology

Q&A per domini ristretti [3]

- Viene utilizzata una **indicizzazione semantica**, per poi poter effettuare una ricerca per keywords, lemmi, synset su indici multipli per i vari metadati.
- Sia la domanda che il **testo** viene **analizzato** sintatticamente e semanticamente (es. “pesca” può essere inteso come sport oppure come un tipo di frutto).
- **Lo scopo** di questo sistema di Q&A è di individuare la porzione di testo che meglio risponde alla domanda formulato. Può anche decidere di non saper rispondere alla domanda.
- Il **sistema di ranking** tiene conto anche di espansioni semantiche dei termini e di allineamenti sintattici e semantici.

Alcuni Esempi



INPS

Informazioni sulla riforma della previdenza complementare ed il trattamento di fine rapporto.

Come si determina l'ammontare del Tfr? Search

LOC DATA ORG PRS MNY MSR NUM

Domanda

Come si determina l'ammontare del Tfr?

Execution time

Question Analysis: 0.203 sec
Atom Retrieval: 1.312 sec
Answer Extraction: 2.344 sec

1	Titolo #10: Come si calcola (full-doc)
* * *	Risposta Il Tfr si determina accantonando, per ciascun anno, una somma pari al 6,91% della retribuzione lorda. Gli importi sono rivalutati, al 31 dicembre di ogni anno, con l'applicazione di un tasso che viene determinato sommando un coefficiente fisso, pari all'1,5%, ed uno variabile, pari al 75% dell'aumento dell'indice dei prezzi al consumo fissato dall'Istat.
2	Titolo #6: La scelta della destinazione del Tfr (full-doc)
* * *	Risposta Le fonti di finanziamento della previdenza complementare per un lavoratore dipendente sono rappresentate da: trattamento di fine rapporto; contributo del lavoratore; contributo del datore di lavoro. La riforma ha previsto che la contribuzione possa avvenire anche soltanto con il versamento del Tfr.
3	Titolo #9: Cos'è il Tfr (full-doc)
* * *	Risposta Il Trattamento di fine rapporto o liquidazione è la somma che viene corrisposta dal datore di lavoro al lavoratore dipendente al termine del rapporto di lavoro, qualunque sia la causa che ne determina la cessazione. Si tratta sostanzialmente di una retribuzione differita nel tempo, che matura di anno in anno in relazione al lavoro prestato e all'ammontare della retribuzione.
4	Titolo #7: La scelta della destinazione del Tfr (full-doc)
* *	Risposta Il lavoratore può scegliere di versare ulteriori contributi, oltre al Tfr. Gli accordi o i contratti collettivi possono stabilire che se il lavoratore decide di contribuire non solo con il Tfr, ma con ulteriori somme, il datore di lavoro è obbligato al versamento di un proprio contributo.

Alcuni Esempi

Question

how do i configure the duration in my device in storing the calendar entries?

Calendar

Calendar settings

Set how long your device stores calendar entries

1. In the calendar, press the Menu key. 2. Click Options. 3. Set the Keep Appointments field. 4. Press the Menu key. 5. Click Save. Your BlackBerry® device deletes calendar entries that are older than the number of days you set. To restore calendar entries, set the Keep Appointments field to a longer period of time. Calendar entries that you received on your BlackBerry device within the set period of time appear in the calendar.

Calendar | Tasks | Contacts | Memos

Calendar settings | Task list settings | Contact list settings | Memo list settings

View the number of contacts, calendar entries, tasks, or memos stored on your device

1. In the address book, calendar, task list, or memo list, press the Menu key. 2. Click Options. The Number of Entries field displays the number of contacts, calendar entries, tasks, or memos that are stored on your device.

Risultati sperimentali: Tester

- **Tester:** è un tool che permette di testare uno specifico servizio e calcolare l'affidabilità del sistema in termini quantitativi.
- **La prima posizione:** per avere un buon sistema di Q&A l'importante mettere la risposta in prima posizione.

JAlgeron - Tester

Service Url

Question List

[TEXT_TARGET]
What do i do to see the mailing list? In the address book, click a mailing list
I cannot pair with another Bluetooth device I cannot pair with a Bluetooth enabled device
Can I do a PIN link? Create a link for a PIN

Answering Time User Time: 2,462 | System Time: 2,292

Question Recall 89,8% (273/304)

Answer Recall 89,9% (275/306)

MRR 0,760

Question-SR @RANK @1: 68,8% | @2: 78,9% | @3: 81,2% | @4: 83,6% | @5: 84,9% | @10: 87,8% | @20: 89,1%

Answer Recall @SCORE @0.01: 88,9% | @0.05: 87,9% | @0.1: 87,9% | @0.2: 81,0% | @0.3: 68,6% | @0.5: 33,7%

Answer Precision @SCORE @0.01: 3,0% | @0.05: 3,4% | @0.1: 4,5% | @0.2: 8,6% | @0.3: 14,6% | @0.5: 47,0%

	Question	User Time	System Time	Answers[RANK]
1	What do i do to see the mailing list?	2,610	2,359	797[1:52]
2	I cannot pair with another Bluetooth device	2,312	2,014	920[1:41]
3	Can I do a PIN link?	1,297	1,624	340[1:27]