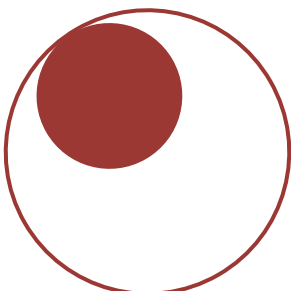
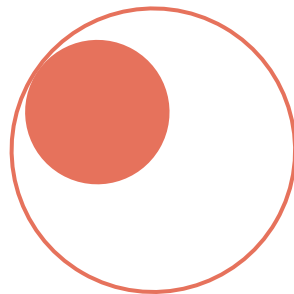
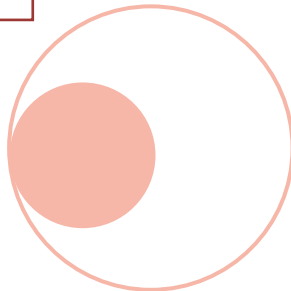
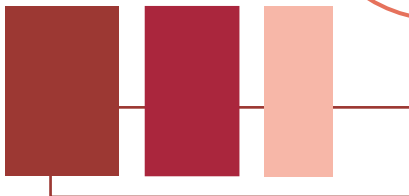


COGITO[®]

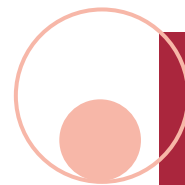


La piattaforma linguistica
di Expert System



 **EXPERT
SYSTEM**
SEMANTIC INTELLIGENCE

COGITO®



COGITO® è la piattaforma linguistica di Expert System, insieme di tecnologie e risorse proprietarie frutto di centinaia di anni/uomo di ricerca e sviluppo.

COGITO® è in grado di gestire in modo intelligente la conoscenza ottimizzando così attività di ricerca, classificazione automatica, estrazione, selezione e correlazione di dati e informazioni.

COGITO®, proprio come farebbe un uomo, interpreta i concetti contenuti nei testi.

Grazie alle tecnologie linguistiche più avanzate, **COGITO®** cattura tutti gli aspetti

strutturali

lessicali

semantici

di un testo, attraverso un'approfondita analisi e disambiguazione di tutti gli elementi che incontra, essendo dotato di una ricca rappresentazione della conoscenza del mondo reale.



COGITO® supera tutti gli altri approcci di trattamento automatico dei documenti e consente di elaborare veramente la conoscenza, dato che non manipola soltanto in modo superficiale un insieme di parole, ma ricorre all'interrogazione di una rete semantica che contiene milioni di informazioni su termini, concetti, abbreviazioni, fraseologie, significati, domini di appartenenza e relazioni tra questi termini.

COGITO® è

- scalabile
- personalizzabile
- trasversale



A proposito di... scalabilità

COGITO® può essere creato per lingue addizionali in modo standard e strutturato grazie alle metodologie e all'esperienza creata per il trasferimento dall'italiano alla seconda lingua, l'inglese.

COGITO® è una piattaforma completa che può essere scalata a tutte le diverse applicazioni per la gestione delle informazioni non strutturate.



A proposito di... personalizzazione

COGITO® è una tecnologia personalizzabile in base ai lessici specialistici dei settori in cui si applica; altrettanto personalizzabili secondo le diverse esigenze sono le componenti: **COGITO®**, infatti, si distingue per il suo carattere di modularità.



A proposito di... trasversalità

COGITO® può essere utilizzato per elaborare contenuti appartenenti ai più diversi domini semantici e dalle strutture differenti: articoli di giornale, documenti tecnici, informazioni finanziarie, manualistica, e-mail, SMS...

Il risultato dell'elaborazione di **COGITO®** è una mappa cognitiva e concettuale del contenuto del testo che consente molteplici tipologie di trattamento automatico dei documenti.

Caratteristiche di COGITO®

COGITO® è dotato di un'architettura pragmatica con caratteristiche che rispondono a precise esigenze espresse dalle aziende. La piattaforma, inoltre, è implementata in modo estremamente efficiente, essendo dotata di un'elevata velocità di esecuzione nel trattamento di grandi volumi di dati e testi.

Inoltre, può essere supportata da diversi ambienti operativi.

Consistenza, affidabilità, robustezza

Nel suo decennale sviluppo, cominciato nel 1992, **COGITO®** con le sue componenti ha conosciuto l'integrazione in prodotti e piattaforme utilizzate da milioni di persone (come lo stesso MicrosoftTM Office), dimostrando doti di grande affidabilità e robustezza.

COGITO® ha sempre dimostrato prestazioni consistenti in tutte le sue implementazioni, complete e personalizzate, sviluppate per le più importanti realtà editoriali, industriali e aziendali.

La garanzia di tale consistenza? Oltre dieci anni di presenza sul mercato e la meticolosa analisi delle singole casistiche, indipendentemente dalla loro peculiarità e complessità.

Elementi necessari all'elaborazione del linguaggio naturale

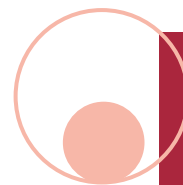
COGITO® si compone di diversi moduli, ognuno dei quali compie un'attività necessaria alla disambiguazione dei testi per l'elaborazione del linguaggio naturale.

L'elaborazione del linguaggio naturale trova un'importante applicazione nella comprensione automatica di domande formulate nel linguaggio di tutti i giorni.

Per comprendere un testo in automatico sono necessari:

- **una rete semantica;**
- **un parser che riconduca ogni testo agli elementi base;**
- **motori linguistici per l'interrogazione della rete semantica (per istituire un confronto tra gli elementi base del testo e la rete semantica dei significati);**
- **un sistema di disambiguazione.**

Componenti di COGITO®



Expert System ha sviluppato i seguenti moduli:

1) SENSIGRAFO ITALIANO

Insieme di reti semantiche strettamente interconnesse a forma di grafo che contengono la rappresentazione concettuale della nostra lingua.



Nella rete sono presenti:

- informazioni relative a relazioni fra oggetti
- specifiche sull'appartenenza ad un dominio lessicale
- informazioni sulla frequenza d'utilizzo.

Più di 2 milioni e 800 mila connessioni: 350.000 tra lemmi, collocazioni e forme fraseologiche, di cui oltre 240.000 synset di sostantivi 50.000 di verbi 60.000 di aggettivi 400.000 collegamenti di iponimia e iperonimia 55.000 collegamenti di iperonimia e troponimia 370.000 collegamenti per il corpus e decine di migliaia di collegamenti di altro tipo (soggetti, oggetti, meronimi).

1 A) SENSIGRAFO GEOGRAFICO

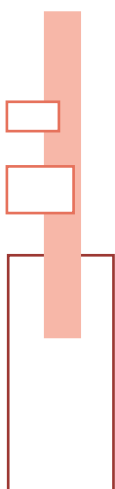
Rappresentazione di tutte le località geografiche significative e rispettivi legami spaziali gestiti in modo intelligente.

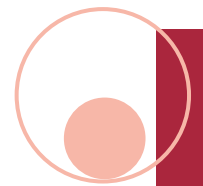
1 B) SENSIGRAFO DI AZIENDE E PRODOTTI

Un sistema che permette di capire quali sono gli oggetti e il concetto di cui si sta parlando. Per distinguere in automatico i nomi comuni dai marchi.

1 C) SENSIGRAFO DI NOMI PROPRI

Oltre 100.000 personaggi, 3.000 società e 20.000 località.





2) PARSER MORFOLOGICO, GRAMMATICALE E SINTATTICO

Il parser esegue una completa analisi morfologica, grammaticale e sintattica della frase in tempi estremamente ridotti, gestendo oltre 3.500 regole. Il parser è costruito ad hoc per l'interrogazione della rete semantica, superando i limiti delle metodologie di parsing esistenti.

3) DIZIONARIO SINONIMI E CONTRARI

Il più completo dizionario dei sinonimi e contrari esistente per la lingua italiana. Inserito in un motore di ricerca permette d'interrogare la base di conoscenza utilizzando la parola ricercata espansa ai suoi sinonimi.

4) CONTROLLO ORTOGRAFICO AVANZATO

La ricchezza dei dizionari sottostanti fanno del controllo ortografico avanzato di Expert System il più accurato esistente in commercio (è l'unico che gestisce in modo approfondito forme flesse, verbi e particelle enclitiche, nomi composti, gestione del contesto per i verbi, ecc.).

5) DIZIONARIO ITALIANO-INGLESE / INGLESE-ITALIANO

Il dizionario contiene oltre 150.000 lemmi e 300.000 fraseologie.

6) DIZIONARI VERTICALI ITALIANI

Raccolta unica di lessici specialistici (informatico, medico, economico...), che consente di analizzare in modo puntuale e completo anche testi specifici di un determinato settore.

7) CLASSIFICATORE AUTOMATICO

Utilizzando sensigrafi e specifiche regole di elaborazione, la tecnologia di classificazione cataloga in modo automatico un documento all'interno di una vasta gerarchia di domini e sottodomini.

8) DISAMBIGUATORE DEI SIGNIFICATI

Per un uomo, il significato è una cosa scontata, perché molti sono gli elementi che vengono in aiuto per capire in quale accezione è utilizzata una parola. Non è così automatico trasferire tale capacità deduttiva ad un programma e ad un sistema di interpretazione e comprensione automatica di un testo. Il disambiguatore dei significati è il programma che analizza singole frasi o interi documenti e distingue il giusto significato per ogni elemento che incontra, eliminando ogni possibile ambiguità.

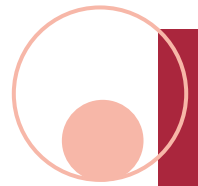
9) TECNOLOGIE DI KNOWLEDGE EXTRACTION

La tecnologia di knowledge extraction di Expert System compie attività di normalizzazione e razionalizzazione di documenti, nonché di estrazione di informazioni significative da testi non strutturati per applicazioni di text mining, come il popolamento automatico di data base.

GESTIONE DI ALTRE LINGUE

Oggi **COGITO**® opera con altre lingue oltre all'italiano, più precisamente è già disponibile la tecnologia per la gestione automatica di testi in lingua **inglese** e in fase di elaborazione per la lingua **francese**.

Analisi linguistica dei testi



La premessa: CORPUS e TAGGING

Il **CORPUS** è una raccolta bilanciata di documenti rappresentativa della totalità degli argomenti che si possono incontrare in qualsiasi tipo di analisi di un testo.

Il **TAGGING** è l'operazione di etichettatura che permette di attribuire delle caratteristiche ad una porzione di testo (tipicamente ad una parola o ad un insieme di parole).



CORPUS e TAGGING rappresentano il 'dietro le quinte' del lavoro di elaborazione automatica del linguaggio, il prerequisito fondamentale di ogni trattamento di testi.

Expert System ha proceduto alla raccolta di un corpus completo e bilanciato in grado di rappresentare in modo significativo la diversa composizione dei tipi possibili di testi esistenti nel mondo reale; ha proceduto poi all'analisi manuale di un sottoinsieme (a sua volta bilanciato) del corpus e alla disambiguazione manuale del significato esatto di tutte le parole polisemiche.

Scopo della raccolta di un corpus

La creazione di tutte quelle regole di disambiguazione necessarie a delegare in misura sempre maggiore a programmi il compito di analizzare e risolvere le ambiguità che si incontrano.

Si traduce in una raccolta, selezione e organizzazione, per ogni significato possibile, delle parole e dei concetti che sono rappresentativi del contesto più comune per ogni significato.

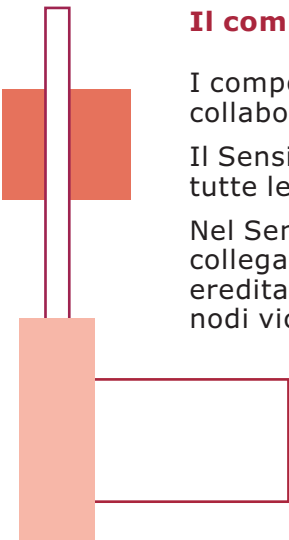
Expert System alimenta continuamente il proprio corpus di riferimento particolarmente ricco e composto di articoli e testi tratti dalle più importanti enciclopedie, quotidiani, riviste, romanzi, saggi, volantini, opuscoli, prestando particolare attenzione ad ottenere un ottimale bilanciamento della conoscenza.

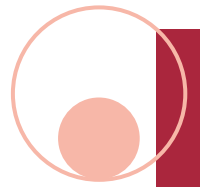
Il comportamento di RETE SEMANTICA e PARSING

I componenti in **COGITO**[®] lavorano sia singolarmente sia in modalità collaborativa.

Il Sensigrafo è la **RETE SEMANTICA** che costituisce la base di conoscenza per tutte le analisi compiute dai motori linguistici costruiti da Expert System.

Nel Sensigrafo ogni synset rappresenta un nodo della rete semantica ed è collegato agli altri da relazioni semantiche in una struttura gerarchica ad ereditarietà. Ogni nodo si arricchisce delle caratteristiche e del significato dei nodi vicini.





È il **PARSER** il motore che identifica i singoli elementi che compongono un testo attribuendogli l'esatta funzione logica e grammaticale.

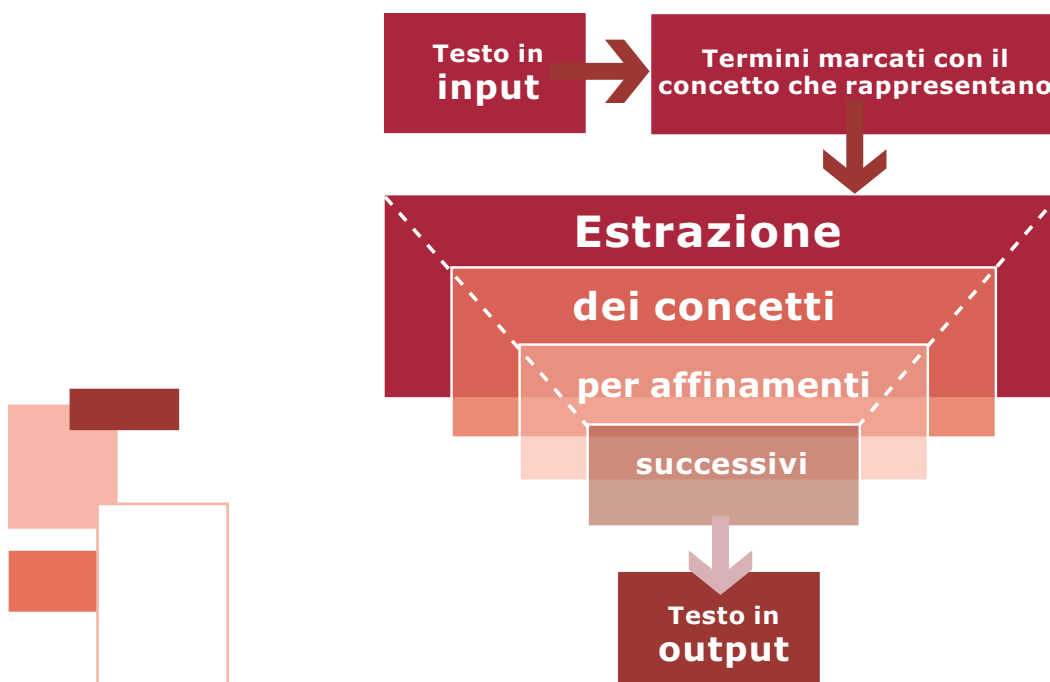
I sistemi attuali mancano di un'integrazione ottimale tra rete semantica e parsing. **COGITO**[®] ha lavorato proprio in quel senso raggiungendo risultati unici.

Passi di analisi:

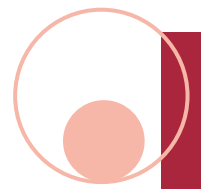
- creazione e mantenimento di una rete semantica di codifica della conoscenza;
- lettura del testo ed estrazione degli elementi delle frasi (motore);
- confronto con la rete di significati;
- estrazione dei concetti dal testo sottoposto ad analisi;
- restituzione del significato (non ambiguo) del testo.

La disambiguazione dei testi

Che cosa significa **disambiguare**?



Disambiguare, dunque...



... significa ricevere testi in input e restituire in output gli stessi testi in cui tutti i termini sono marcati con il concetto che rappresentano. Dal punto di vista computazionale, si traduce in una serie di passaggi di analisi del testo e di affinamento dell'interpretazione dei concetti contenuti in esso.

Perché ad un programma bisogna trasmettere un'interpretazione del mondo univoca costruendogli un sistema di riferimento che sia l'equivalente dell'esperienza del mondo dell'uomo: un'esperienza generica, chiaramente, che prescindia dall'individualità.

Il linguaggio deve tornare ad essere un puro veicolo di significato.

Passi nello specifico della disambiguazione

Per analizzare la frase:

"Il commesso dice che non accetta la carta di credito"

si compirà una sequenza di analisi che possiamo descrivere per passi.



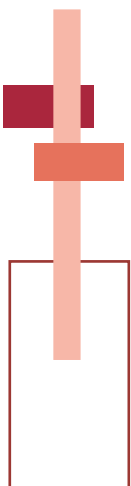
Primo passo

Analisi morfologica e grammaticale:

articolo **IL** determinativo, maschile, singolare
sostantivo **COMMESSE** maschile, singolare
verbo **DIRE** terza persona singolare, indicativo presente, transitivo, attivo
congiunzione **CHE**
avverbio **NON**
verbo **ACCETTARE** terza persona singolare, indicativo presente, transitivo, attivo
articolo **LA** determinativo, femminile, singolare
sostantivo **CARTA** femminile, singolare
preposizione **DI** invariabile in genere, invariabile in numero
sostantivo **CREDITO** maschile, singolare

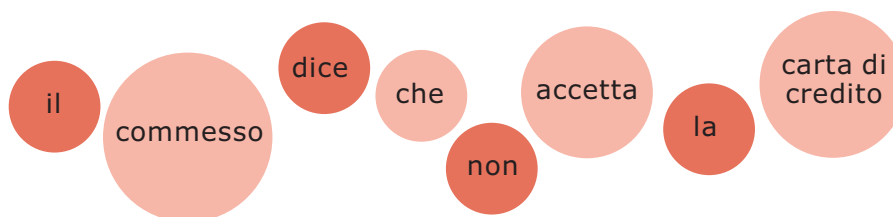


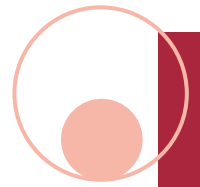
La frase si trasforma da mera sequenza di caratteri a insieme di termini grammaticalmente valorizzati.



Secondo passo

Raccolta di collocazioni/locuzioni (tramite le informazioni del Sensigrafo)





L'analisi compiuta dal Sensigrafo

Si stabilisce la priorità da dare alle singole parole o a gruppi di due o più parole quando queste diventano funzionanti come unità grammaticale e lessicale. *Carta* e *credito* sono parole dotate di significato proprio, ma in questo caso devono essere raccolte all'interno dell'espressione *carta di credito*.

Bilanciamento della conoscenza del mondo e conoscenza linguistica di base molto vasta

Permette di generalizzare questo tipo di analisi per tutta la lingua.

Terzo passo

Analisi logica e sintattica:

il commesso (Soggetto) [Inizio frase reggente]
dice (Predicato verbale) [Fine frase reggente]
che [Inizio frase subordinata]
non accetta (Predicato verbale)
la carta di credito (Complemento oggetto) [Fine frase subordinata]

Arrivati a questo terzo passaggio, tutti i sistemi di analisi attualmente esistenti si fermano e non assegnano ai diversi termini componenti la frase il giusto significato, rendendo impossibile un'analisi concettuale vera e propria.

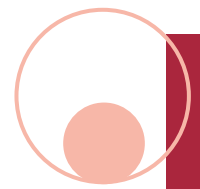
Per procedere alla disambiguazione vengono evidenziati i termini da disambiguare:

commesso
dire
accettare
carta di credito

Ognuno dei significati (**synset**) relativi ad ogni coppia lemma/ruolo grammaticale ottiene una percentuale di probabilità in base a:

- frequenza d'uso
- domini (1)
- attributi su aggettivi/sostantivi (2) e controlli di congruenza semantica
- attributi sui verbi (3)
- informazioni contestuali, fondamentali per risolvere casi di ambiguità complessa

Alla fine viene considerato come assegnato il synset che ottiene la percentuale più alta.



Controlli di congruenza semantica

Per quanto riguarda i controlli di congruenza semantica, i programmi di Expert System arrivano a rispondere correttamente a domande come: "Il gatto si nutre" tramite regole di derivazione autogenerate in base alla propria rappresentazione del mondo.

Dall'oggetto "Animale" nel Sensigrafo derivano gli oggetti "Vertebrato", "Invertebrato" e "Animali Marini".

Regola numero 1:

Ogni oggetto derivato eredita tutte le proprietà dell'oggetto da cui viene derivato. Proprietà di ereditarietà.

Continuano le derivazioni
"Animale" ---> "Mammifero terrestre" --->
"Carnivoro" ---> "Felide" ---> "Felino" ---> "Gatto"
---> "Persiano" ---> ...

Elaborazioni:

Vive (Animale) = Vero
Vive (Mammifero) = Vero
Si Nutre (Animale) = Vero
Si Nutre (Mammifero) = Vero
Si Nutre (Carnivoro) = Vero

Allatta (Mammifero) = Vero
Allatta (Felide) = Vero
Allatta (Felino) = Vero
Miagola (Gatto) = Vero

**Risposta:
Sì, il gatto si nutre.**

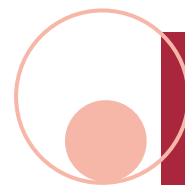
Problematiche

La disambiguazione del significato è uno dei problemi più complessi affrontati dalla linguistica.

Per restare in tempi di elaborazione accettabili sono necessari:

- una conoscenza di tipo enciclopedico;
- un insieme di algoritmi di disambiguazione perfettamente ottimizzati.

focus on: il Disambiguatore



Disambiguare: ecco il vero problema dell'interpretazione dei testi in automatico. Perché un programma possa distinguere

**“La ruggine mangia la torre”
da “Il cavallo mangia la torre”**

bisogna "insegnargli" che in ogni lingua esistono numerosissime ambiguità che un uomo risolve senza problemi.

Ma quello che un uomo sa per istruzione o esperienza, un software deve dedurlo dal testo in automatico, grazie a conoscenza codificata e ad avanzate tecnologie.

La ricerca e sviluppo di sistemi automatici di disambiguazione linguistica deve risolvere un problema cruciale: la gestione del **numero di combinazioni possibili** che si possono generare quando si tratta con parole e testi. Queste possono combinarsi insieme in un numero di modi elevatissimo, con crescita esponenziale.

La problematica che più di frequente si trovano a dover affrontare i motori di disambiguazione è proprio quella di ridurre drasticamente il numero di possibilità, dato che oltre certi valori di soglia, l'aumento di precisione anche di pochi punti percentuali (quando possibile) ha costi elevatissimi (di tempo e ricerca).

Un altro limite delle soluzioni esistenti, è quello di non essere estendibili da singoli campioni all'universo senza ottenere percentuali di errore anche molto elevate; i sistemi diventano velocemente instabili e non sono in grado di risolvere il problema della disambiguazione in ambito linguistico per ogni lingua e/o dominio di conoscenza.

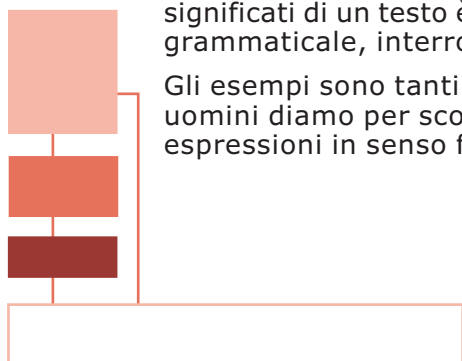
Il disambiguatore di COGITO® è in grado di compiere “ragionamenti” che distinguono i diversi significati di tutti gli elementi di un testo individuando il contesto in cui questi sono collocati.

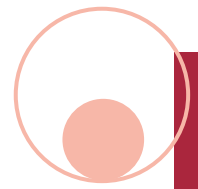
La difficoltà di creazione di un sistema in grado di lavorare su tutte le possibili ambiguità di una lingua è dimostrata dagli scarsi risultati di mercato raggiunti dalla maggior parte dei sistemi esistenti.

Il disambiguatore sviluppato da Expert System, dunque, ha carattere di eccezionalità ed è il risultato più ambizioso che si prepone la linguistica: far comprendere ad un programma i concetti delle frasi in modo simile a come li intenderebbe un uomo grazie alla sua esperienza e cultura.

Il disambiguatore può lavorare per frasi o per documenti, a seconda di come opera la scansione delle frasi. Quella del discernimento di tutti i possibili significati di un testo è un'attività che va ad aggiungersi alle altre analisi: logica, grammaticale, interrogazione del Sensigrafo, analisi dei domini.

Gli esempi sono tanti: tanti quante sono le interpretazioni delle parole che noi uomini diamo per scontate ma che non lo sono per un programma, comprese espressioni in senso figurato.





Alcuni esempi di quello che può fare il disambiguatore?

Comprendere univocamente il significato delle frasi che seguono.

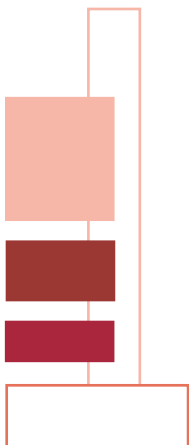


E che dire di:

Maria **ha partorito** un bel bambino

Maria **ha partorito** un bel progetto

Il disambiguatore suggerisce che nel primo caso si parla di partorire come mettere al mondo, nel secondo caso come produrre sentimenti e idee. Intelligente come un uomo.



Confronto con le altre tecnologie

Il sistema attualmente più utilizzato per la comprensione e l'elaborazione di testo scritto nel linguaggio di tutti i giorni è quello conosciuto con il nome di **backtracking**, che dà la possibilità di tornare indietro sugli step precedenti e ripercorrere strade già percorse qualora, procedendo con l'analisi, si verificassero eventi che mutano le condizioni e la validità delle proprie affermazioni precedenti.

Questo metodo è generico e scalabile ma soffre più di qualunque altro approccio del problema dell'esplosione esponenziale dell'albero delle combinazioni possibili: la sfida è quella di conservare l'idea di base (cioè la possibilità di interazione fra diversi scenari possibili al fine di scegliere quello migliore) riducendo di diversi ordini di grandezza i tempi di analisi (un sistema che impieghi un'ora a disambiguare un breve documento non serve a nessuno).

Precisione di disambiguazione

- ➔ i sistemi attuali, per contesti liberi, hanno un livello di disambiguazione che va dal 45% al 55% riducendo il contesto, alcuni sistemi raggiungono un livello che non supera il 73-75%;
- ➔ riducendo ancora il numero di significati di riferimento (con una grossolana interpretazione dei significati) raggiungono punte attorno all'80% (anche se si tratta di prototipi non validi per applicazioni reali);
- ➔ **Expert System grazie alla propria tecnologia in contesti liberi ottiene una precisione superiore al 90%.**

Expert System S.p.A.

Via Virgilio 56/Q
41100 Modena - Italy

tel: +39 059 894011
fax: +39 059 894099

Via Machiavelli 47
00185 Roma - Italy

www.expertsystem.it
info@expertsystem.it

Expert System S.p.A. è leader di mercato nella realizzazione di soluzioni avanzate di **Semantic Intelligence** per aziende ed enti governativi.

Unica realtà italiana a fornire a Microsoft tecnologie avanzate integrate in tutti i suoi principali prodotti, Expert System fa della **gestione "intelligente" delle informazioni non strutturate** il proprio core business.