

tesi di laurea

Sperimentazione di un processo per l'interoperabilità semantica

Anno Accademico 2008 - 2009

relatore

Ch.mo prof. Angelo Chianese

Ch.mo prof. Porfirio Tramontana

candidato

Roberto De Caro

Matr. 534/2836

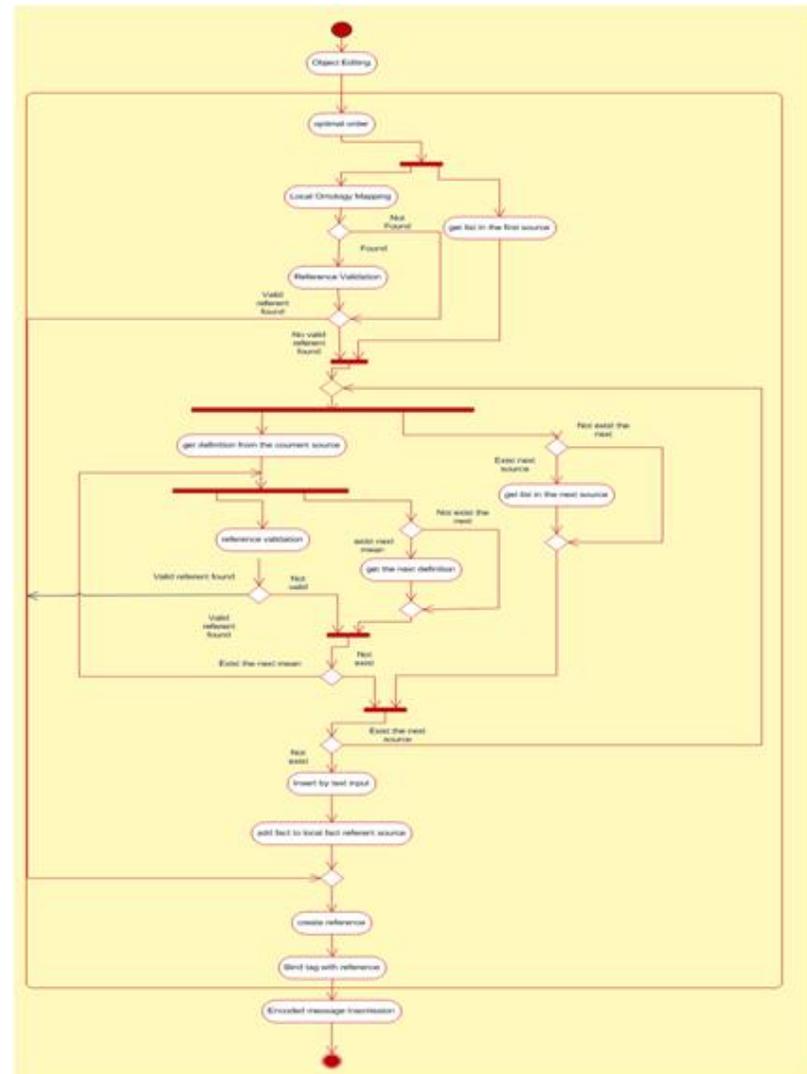
Processo ottimizzato

❑ L'obiettivo è quello di abbinare ad ogni tag un riferimento ad un concetto presente in una ontologia locale o globale.

❑ I significati da associare a tali tag vengono ricercati nelle varie fonti automaticamente.

❑ Se nessun significato è valido o corretto, si fa un inserimento manuale.

Sperimentazione di un processo per l'interoperabilità semantica



Sperimentazione

Questa sezione illustra un esperimento che è stato effettuato per valutare i punti di forza e punti deboli del modello di comunicazione proposto. Per raggiungere questo obiettivo, è stato considerato uno scenario del processo di comunicazione e la qualità dei risultati ottenuti.

Obiettivi

- ❖ Evidenziare come variano i tempi di accesso al variare delle variabili in gioco

Variabili in gioco

- ❖ Grado di riempimento del database locale
- ❖ Ordine di accesso alle fonti
- ❖ Tipo di persone
- ❖ Tipo di tag

Protocollo Sperimentale

- ✓ Fonti: pagine riguardanti tre domini: sport, musica e film
- ✓ Raccolta 200 documenti
- ✓ Estrazione di 500 tag
- ✓ Sottomissione casuale al sistema

❖ Collegandoci a tale indirizzo:

http://143.225.229.92:8080/WS_Client, e con apposite credenziali potremo avviare l'applicazione

❖ Associazione di referenti e referenze ai tag

❖ Costruzione dell'ontologia locale

❖ Ricerca su più domini

❖ Inserimento di nuovi referenti

Protocollo sperimentale



IO Parameters

Source: Sfoglia...
(htm/doc)

Tag List: Sfoglia...

Log-In

Username:

Access Key:

Fields with (*) must be filled

Start

Sperimentazione: metriche

AVC: media del numero di accessi a possibili significati di un tag necessari al processo di codifica

LIR: numero di creazioni / numero totale di tag

Tval: tempo di validazione, è il tempo che occorre per validare un singolo tag

Tcre: tempo di creazione, è il tempo che occorre per immettere un significato manualmente.

α : Tval / numero accessi

β : Tcre / numero di creazioni

Effort: sforzo o costo dell'applicazione espresso in termini di tval e tcre

EFFORT totale = α medio * numero di accessi + β medio * numero di creazioni

Esempio guida

- **Sperimentazione del processo**
- **Tipico esempio per capire come viene fatta la sperimentazione**
- **Vengono processati vari tag e calcolate le metriche necessarie**
- **Due modi: DB=0 e DB<>0 con tag sportivi e tutti i tag**

| Numero di tag | AVC | M (tval) | σ^2 (AVC) | σ (AVC) | σ^2 (tval) | σ (tval) |
|---------------|------|----------|------------------|----------------|-------------------|-----------------|
| 42 | 3,12 | 24,6 | 19,15 | 4,38 | 464,79 | 21,56 |

Sperimentazione di un processo per l'interoperabilità semantica

| | | | Numero di accessi | LIR | EFFORT | | |
|--------------|-------------|-------|-------------------|-------|--------|------|----|
| | | | | | tval | tcrc | |
| Sport | Juventus | DB=0 | 1 | 0 | 9 | | |
| | Chelsea | | 1 | | 9 | | |
| | Gattuso | | 1 | | 9 | | |
| | Rome | | 1 | 0 | 12 | | |
| | Rossi | | 2 | | 19 | | |
| | Lorenzo | | 21 | | 133 | | |
| | Contador | | 1 | | 20 | | |
| | Ferrara | | 2 | 2 | 31 | 11 | |
| | Napoli | | 2 | | 19 | | |
| | Santacroce | | 0 | | 17 | 7 | |
| | Italy | | 1 | | 11 | | |
| | Bianconeri | | 1 | | 13 | | |
| | Tutti | | Ronaldo | DB=0 | 1 | 0 | 11 |
| Messi | | 1 | 13 | | | | |
| Goal | | 3 | 21 | | | | |
| Players | | 2 | 0 | | 23 | | |
| Herbie | | 2 | | | 39 | | |
| Batman | | 1 | | | 14 | | |
| xXx | | 1 | | | 10 | | |
| Cry | | 8 | 0 | | 46 | | |
| Eminem | | 1 | | | 14 | | |
| Nelly | | 1 | | | 11 | | |
| Transamerica | | 12 | | | 2 | 33 | |
| Capote | | 5 | | | | 24 | 12 |
| Firewall | | 4 | | | 39 | 8 | |
| Sport | AnceIotti | DB<>0 | 1 | 0 | 16 | | |
| | Benitez | | 1 | | 14 | | |
| | Liverpool | | 1 | | 13 | | |
| | Ranieri | | 1 | 2 | 23 | 11 | |
| | Lippi | | 3 | | 29 | 8 | |
| | Alonso | | 2 | 3 | 21 | 9 | |
| | Ferrari | | 4 | | 35 | 8 | |
| | Montezemolo | | 1 | | 24 | 8 | |
| | Tutti | | Eminem | DB<>0 | 14 | 1 | 14 |
| Madonna | | 2 | 9 | | | | |
| Pink | | 15 | 78 | | 6 | | |
| Terry | | 3 | 1 | | 39 | | |
| Chelsea | | 1 | | | 9 | | |
| Drogba | | 1 | | | 15 | | |
| Calderon | | 3 | | | 32 | 11 | |
| 2012 | | 1 | 0 | | 26 | | |
| Firewall | | 1 | | | 39 | | |

Sperimentazione: risultati

| | Media Alpha | Media Beta | Varianza Alpha | Varianza Beta | Deviazione standard Alpha | Deviazione Standard Beta |
|-------------|-------------|------------|----------------|---------------|---------------------------|--------------------------|
| Tutti i tag | 12,99 | 47,10 | 167,75 | 9111,02 | 12,95 | 95,45 |

Facendo la sperimentazione con un gran numero di tag otteniamo che la $M(\text{Beta}) > M(\text{Alfa})$, ciò è ovvio.

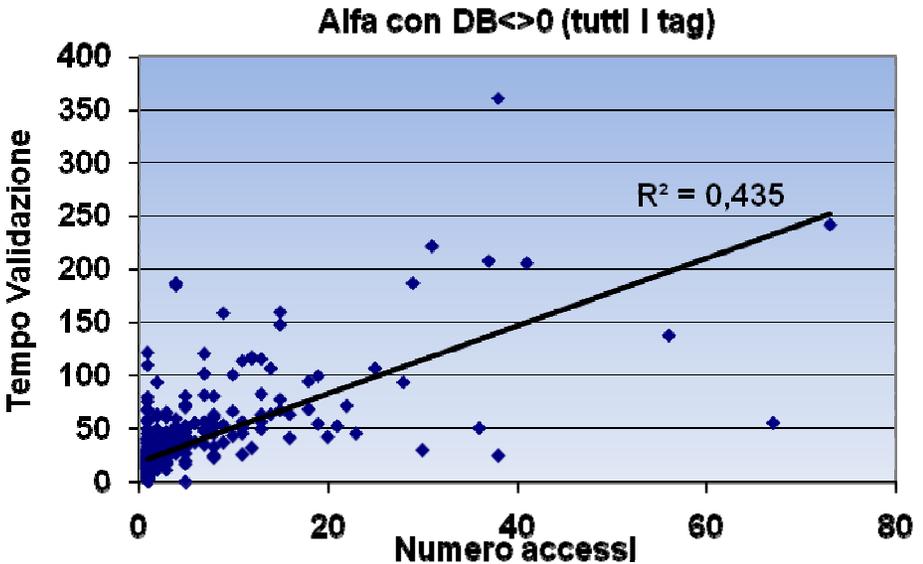
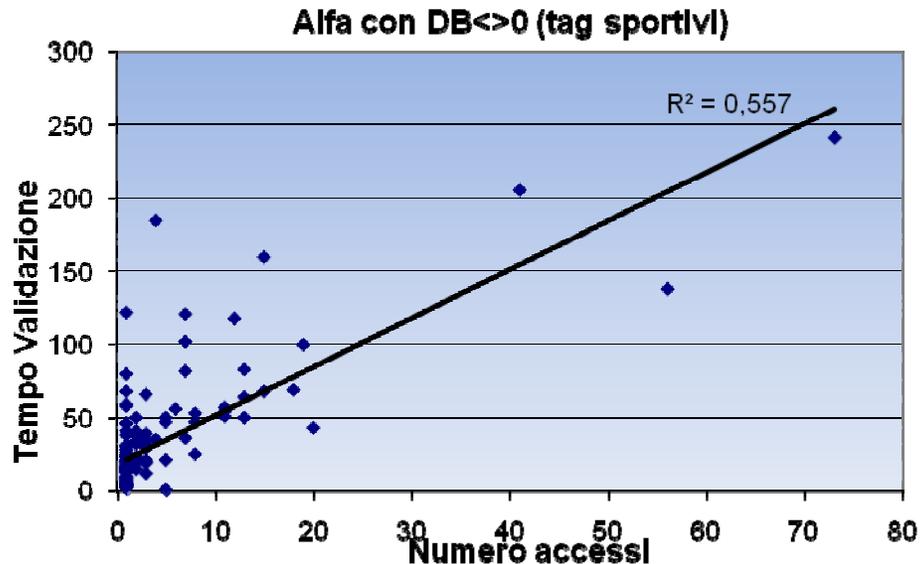
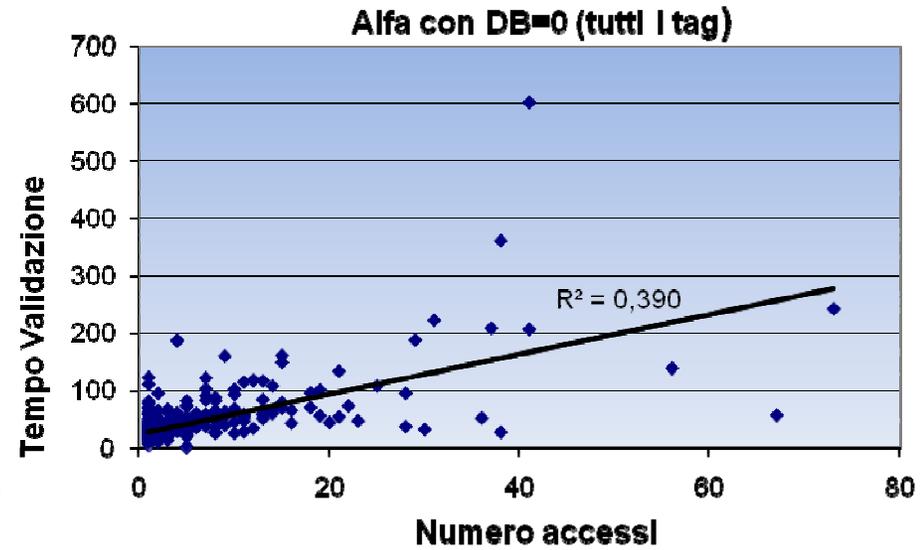
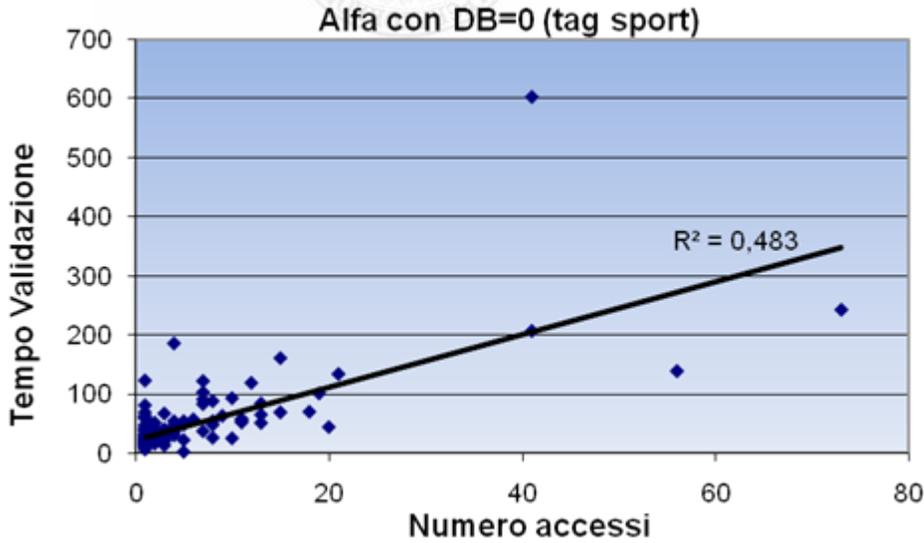
La varianza sia di alpha che di beta è abbastanza distante dallo zero, cioè dalla situazione ideale.

Sperimentazione: risultati

| | Numero tag | Media tval | Deviazione standard tval | Alpha medio | Deviazione standard alpha |
|------------------|------------|------------|--------------------------|-------------|---------------------------|
| Tag Sportivi | 144 | 44,38 | 60,48 | 17,51 | 15,74 |
| Tutti i tag DB=0 | 362 | 42,61 | 49,01 | 17,09 | 18,04 |

| | Numero tag | Media tval | Deviazione standard tval | Alpha medio | Deviazione standard alpha |
|--------------------|------------|------------|--------------------------|-------------|---------------------------|
| Tag sportivi DB<>0 | 139 | 32,06 | 40,01 | 4,09 | 8,99 |
| Tutti i tag DB<>0 | 409 | 32,90 | 39,25 | 13,17 | 13,61 |

Sperimentazione di un processo per l'interoperabilità semantica



Conclusioni e Sviluppi futuri

- **Minacce degli esperimenti valutate**
 - **Dipendenza dall'omogeneità delle fonti**
 - **Dipendenza dal riempimento del database**
 - **Si è cercato di valutare il rapporto tra lo sforzo totale e il numero di validazione e creazione da fare**

- **Il tempo di accesso alle fonti proveniente da altre tesi si è rilevato non tanto pesante dei tempi di validazione e di creazione**

- **Possibili sviluppi**
 - **Valutazione dei risultati della sperimentazione effettuata per scenari differenti e con testing set differenti ed eterogenei**
 - **Valutazione dalla dipendenza dello sperimentatore**
 - **Valutazione dalla maggior quantità dei dati**