

ELN 2007

Cross-Language Evaluation Forum

Multilingual Textual Document Retrieval (Ad Hoc)

Professor: Amedeo Cappelli
Studente: Camaroto Alessio
Matricola: 251823
E-mail: camaroto@cli.di.unipi.it



CLEF

- ☐ CLEF supporta applicazioni globali per biblioteche digitali
 - ☐ Sviluppando una infrastruttura per
 - ☐ Effettuare prove
 - ☐ Valutare operato
 - ☐ dei sistemi di IR basati su lingue europee in contesti
 - ☐ Monolingua
 - ☐ Cross language
- ☐ Creando suite di dati riutilizzabili dagli sviluppatori di sistemi per operazioni di benchmark

CLEF - Scopi

- ☐ Creare una comunità di ricercatori e distributori
 - ☐ Studiare gli stessi problemi
 - ☐ Facilitare collaborazioni future tra gruppi con stessi interessi
 - ☐ Stabilire collegamenti forti, scambio di idee e condivisione risultati in US, Asia ed Europa
- ☐ Assistere e stimolare lo sviluppo di sistemi di CLIR basati su lingue europee

CLEF - Homepage

CLEF 2006 Workshop, 20-22 September, Alicante, Spain

- CLEF 2006 ad hoc track si è occupata di soluzioni ad operazioni
 - **Monolingua**, richiesta e ricerca di documenti nella stessa lingua
 - **Bilingue**, richiesta in una lingua e risultati in un'altra
 - **Robust**, soluzioni alla ricerca di documenti con query molto complicate (pensato per partecipanti più esperti)
 - Monolingua, bilingue e Multilingua

Monolingua e bilingue

- Operazioni Monolingua
 - Bulgaro, Francese, Ungherese, Portoghese
- Operazioni bilingue, con X qualsiasi
 - X-> Bulgaro
 - X-> Francese
 - X-> Ungherese
 - X-> Portoghese

Robust

- Studiare e ottenere buone prestazioni per
 - Richieste "difficili"
 - Richieste mal risolte negli anni precedenti
- Operazioni Monolingua, bilingue e Multilingua
 - Su testi raccolti tra il 2001 e il 2003, in 6 lingue
 - Tedesco, Inglese, Francese, Olandese, Italiano e Spagnolo

Track Setup (1)

- CLEF adotta un corpus, con un metodo automatico di valutazione delle prestazioni del sistema
- La test collection usata è formata da
 - Un insieme di topics, descrivono le informazioni necessarie contenute nei documenti
 - Una collezione di documenti, su cui ricercare quelli che soddisfano le informazioni precedenti

Track Setup (2)

- La valutazione dei documenti viene fatta considerando
 - Relevance
 - Recall
 - Precisionrispetto ai documenti trovati, fissando un topic ben preciso
- Questo paradigma è stato esportato in contesti multilingua per la creazione di test collection

Test Collections

- Documenti**, parte del corpus multilingua del CLEF realizzato con
 - Quotidiani
 - News delle agenzie
- Topics**, strutturati in una serie di dichiarazioni che rappresentano le informazioni necessarie
 - Usati dai sistemi per definire le loro richieste

Documenti (1)

- Monolingua e bilingue**
 - Inglese, quotidiani americani e inglesi (94-95)
 - Francese, quotidiani francesi più agenzie svizzere (94-95)
 - Portoghese, quotidiani portoghesi e brasiliani (94-95)
 - Bulgaro, quotidiani (02)
 - Ungherese, quotidiani (02)

Table 1. Test collections for the main stream Ad Hoc tasks.

| Language | Collections |
|------------|---------------------------------|
| Bulgarian | Sega 2002, Standart 2002 |
| English | LA Times 94, Glasgow Herald 95 |
| French | ATS (SDA) 94/95, Le Monde 94/95 |
| Hungarian | Magyar Hirlap 2002 |
| Portuguese | Público 94/95; Folha 94/95 |

Documenti (2)

- Robust**, circa 1.35 milioni di documenti (3.6 GB di testo)
 - Inglese, Tedesco, Olandese, Francese, Italiano e Spagnolo
- Usati al CLEF 2001, 2002 e 2003

Table 2. Test collections for the Robust task.

| Language | Collections |
|----------|--|
| English | LA Times 94, Glasgow Herald 95 |
| French | ATS (SDA) 94/95, Le Monde 94 |
| Italian | La Stampa 94, AGZ (SDA) 94/95 |
| Dutch | NRC Handelsblad 94/95, Algemeen Dagblad 94/95 |
| German | Frankfurter Rundschau 94/95, Spiegel 94/95, SDA 94 |
| Spanish | EFE 94/95 |

Topics (1)

- Formato da 3 parti
 - "Title", breve e significativo
 - "Description", di una sola frase
 - "Narrative", descrizione più complessa che specifica i criteri di valutazione delle prestazioni

```
<numero topic> <lingua> <title> <description>
<top> <num> C302 </num>
<EN-title> Consumer Boycotts </EN-title>
<EN-desc> Find documents that describe or discuss the impact of consumer boycotts. </EN-desc>
<EN-narr> Relevant documents will report discussions or points of view on the efficacy of consumer boycotts. The moral issues involved in such boycotts are also of relevance. Only consumer boycotts are relevant, political boycotts must be ignored. </EN-narr> </top>
<narrative>
```

©camaroto@cli.di.unipi.it Multilingual Textual Document Retrieval (Ad Hoc) 12/07/2007 13

Topics (2)

- Monolingua e bilingue**, realizzati 75 topics
 - Per rendere più semplice la valutazione dei risultati su differenti collezioni
 - Impossibile costruire insiemi di topics per documenti di periodi di tempo diversi (1994-95 e 2002)
- Suddivisione**
 - 25 comuni ai documenti del 1994-95 e del 2002
 - 25 speciali per Inglese, Francese e Portoghese (1994-1995)
 - 25 speciali per Bulgaro e Ungherese (2002)

©camaroto@cli.di.unipi.it Multilingual Textual Document Retrieval (Ad Hoc) 12/07/2007 14

Topics (3)

- Preparati in diverse lingue
 - Europee
 - Bulgaro, Inglese, Francese, Tedesco, Ungherese, Italiano, Portoghese e Spagnolo
 - Non Europee
 - Aramaico, Cinese, Hindi, Indonesiano, Oromo e Telegu

©camaroto@cli.di.unipi.it Multilingual Textual Document Retrieval (Ad Hoc) 12/07/2007 15

Topics (4)

- Robust**, 160 topics
 - 60 usati per provare i sistemi
 - 100 usati per le valutazioni
- Basati sui topics dei CLEF 2001, 2002 e 2003
- Disponibili per
 - Inglese, Tedesco, Francese, Italiano, Spagnolo e Olandese

©camaroto@cli.di.unipi.it Multilingual Textual Document Retrieval (Ad Hoc) 12/07/2007 16

Come Partecipare

- Ogni gruppo deve
 - Iscrivarsi al CLEF
 - Inviare almeno un title+description ("TD") per prova
 - In modo da aumentare la compatibilità tra gli esperimenti
 - Testare il proprio sistema sui documenti della collezione del CLEF

Assegnamento della Rilevanza

- Essendo le test collections molto grandi è impossibile calcolare la rilevanza di ogni documento
- Ogni gruppo invia i propri risultati al CLEF
 - Raccolti e divisi per ogni topic e lingua
 - Ordinati in base alla frequenza di invio
 - Presi quelli con la frequenza maggiore
 - Sono quelli di cui si calcola la rilevanza e usati per fornire le valutazioni dei sistemi

Calcolo dei Risultati

- L'Efficacia dei sistemi viene calcolata in base
 - Risultati forniti dai partecipanti
 - Rilevanza dei documenti
- Monolingua e bilingue
 - Recall
 - Precision
 - Mean Average Precision (MAP)
- Robust
 - Geometric Average Precision (GMAP)
 - http://trec.nist.gov/trec_eval/trec_eval.8.0.tar.gz

Recall

- % di documenti rilevanti recuperati dal sistema

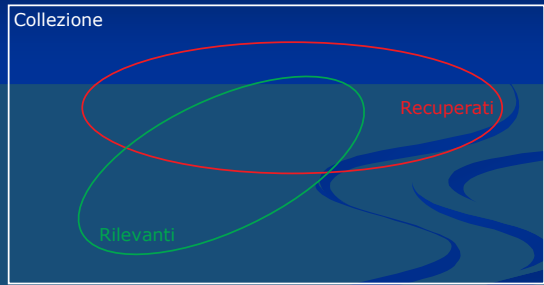
$$\text{recall} = \frac{\text{numero di documenti rilevanti trovati}}{\text{numero di documenti rilevanti della collezione}}$$

Precision

▫ % di documenti recuperati, dal sistema, rilevanti

$$\text{precision} = \frac{\text{numero di documenti rilevanti trovati}}{\text{numero totale dei documenti ritrovati}}$$

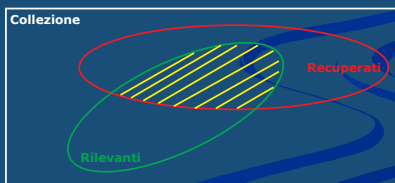
Relevance



Precision vs. Recall

▫ **Precision** = % di documenti recuperati che sono rilevanti

▫ **Recall** = % di documenti rilevanti che sono stati recuperati



Come Calcolarli

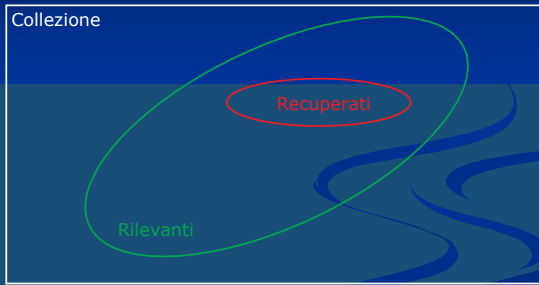
▫ **Precision** = frazione di documenti recuperati che sono rilevanti

▫ **Recall** = frazione di documenti rilevanti che sono stati recuperati

| | Rilevanti | Non Rilevanti |
|----------------|----------------------------|----------------------------|
| Recuperati | tp (true positive) | fp (false positive) |
| Non recuperati | fn (false negative) | tn (true negative) |

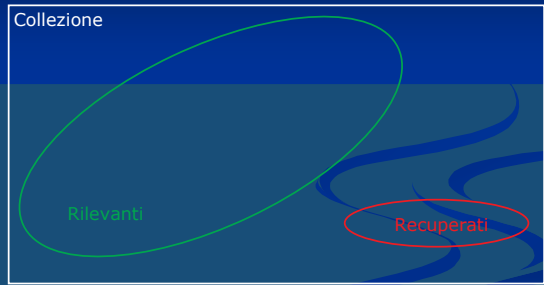
$$\text{Precision } P = \frac{tp}{(tp + fp)}$$
$$\text{Recall } R = \frac{tp}{(tp + fn)}$$

1° Caso



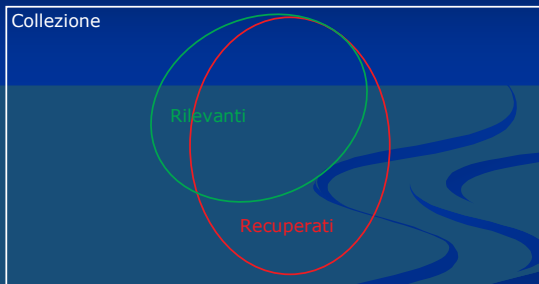
Alta precision, recall molto bassa

2° Caso



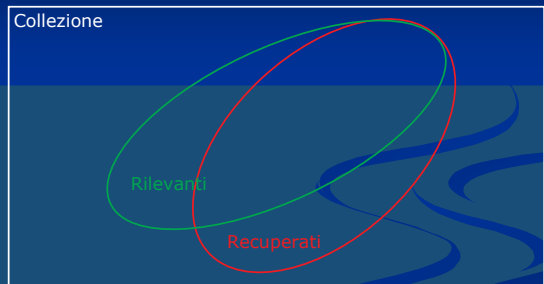
Precision e recall molto basse

3° Caso



Bassa precision, recall molto alta

4° Caso



Precision e recall molto alte

Average Precision (AP)

- Media della precisione di un documento rilevante trovato
- Calcolata di solito sul 1000 ricerche positive

$$AP = \frac{\sum_i \text{precision}(i)}{1000} \quad 1 \leq i \leq 1000$$

Mean Average Precision (MAP)

- Calcolo della precisione media per ogni query (AP)
- Calcolo della media matematica delle AP su tutte le queries effettuate

$$MAP = \frac{\sum_i AP(i)}{\#queries} \quad 1 \leq i \leq \#queries$$

Geometric Average Precision (GMAP)

- Calcolo della precisione media per ogni query (AP)
- Calcolo della media geometrica delle AP su tutte le queries effettuate

$$GMAP = \sqrt[n]{\prod_n AP_n} = \exp \frac{1}{n} \sum_n \log AP_n \quad n=50 \text{ TREC}$$

Classifiche (1)

Table 6. Best entries for the monolingual track.

| Track | Participant Rank | | | | | Diff. |
|------------|------------------|---------------|-------------|--------------|-------------|----------------------|
| | 1st | 2nd | 3rd | 4th | 5th | |
| Bulgarian | unine | rsi-jhu | hummingbird | daedalus | | 1st vs 4th 20.90% |
| | 33.14% | 31.98% | 30.47% | 27.87% | | |
| | UniNEbg2 | 02aplmobgtdd | humBG06tde | bgFSbg2S | | |
| French | unine | rsi-jhu | hummingbird | alicante | daedalus | 1st vs 5th 17.76% |
| | 44.68% | 40.96% | 40.77% | 38.28% | 37.94% | |
| | UniNEfr3 | 95aplmofrt5sl | humFR06tde | 8dfrexp | frFSfr2S | |
| Hungarian | unine | rsi-jhu | alicante | mokk | hummingbird | 1st vs 5th 28.26% |
| | 41.35% | 39.11% | 35.32% | 34.95% | 32.24% | |
| | UniNEhu2 | 02aplmohutdd | 30dfrexp | plain2 | humHU06tde | |
| Portuguese | unine | hummingbird | alicante | rsi-jhu | u.buffalo | 1st vs 5th 12.31% |
| | 45.52% | 45.07% | 43.08% | 42.42% | 40.53% | |
| | UniNEpt1 | humPT06tde | 30okapiexp | 95aplmopttd5 | UBptTDrfl | |

Classifiche (2)

Table 7. Best entries for the bilingual task.

| Track | Participant Rank | | | | | Diff. |
|-------------------|------------------|--------------|--------------|--------------------|--------------------|------------|
| | 1st | 2nd | 3rd | 4th | 5th | |
| Bulgarian | daedalus | | | | | |
| MAP | 17.39% | | | | | |
| Run | beFSbzWen2S | | | | | |
| French | umne | queenmary | rsi-jhu | daedalus | | 1st vs 4th |
| MAP | 41.92% | 33.96% | 33.60% | 33.20% | | 26.27% |
| Run | UniNEBfr1 | QMUL06e2f10b | aplbienfrd | frFSfrSen2S | | |
| Hungarian | daedalus | | | | | |
| MAP | 21.97% | | | | | |
| Run | hmfShuMen2S | | | | | |
| Portuguese | umne | rsi-jhu | queenmary | u.buffalo | daedalus | 1st vs 5th |
| MAP | 41.38% | 35.49% | 35.26% | 29.08% | 26.50% | 55.85% |
| Run | UniNEBpt2 | aplbiesptd | QMUL06e2p10b | UBen2pTDrft2 | ptFSptSen2S | |
| English | rsi-jhu | depok | lirc | celi | dsv | 1st vs 5th |
| MAP | 32.57% | 26.71% | 25.04% | 23.97% | 22.78% | 42.98% |
| Run | aplbimen5 | Ulj_d_int | OMTD | CELLtohoNEXPANSHON | DivAhhbAgPulldofmz | |

Classifiche (3)

Table 8. Best entries for the robust monolingual task.

| Track | Participant Rank | | | | | Diff. |
|----------------|------------------|---------------|----------------|----------------|--------------|------------|
| | 1st | 2nd | 3rd | 4th | 5th | |
| Dutch | hummingbird | daedalus | colesir | | | 1st vs 3rd |
| MAP | 51.06% | 42.39% | 41.60% | | | 22.74% |
| GMAP | 25.76% | 17.57% | 16.40% | | | 57.13% |
| Run | humNL00trdo | nlFSnR2S | CoLesIRnITst | | | |
| English | hummingbird | reina | dcu | daedalus | colesir | 1st vs 5th |
| MAP | 47.63% | 43.66% | 43.48% | 39.69% | 37.64% | 26.54% |
| GMAP | 11.69% | 10.53% | 10.11% | 8.93% | 8.41% | 39.00% |
| Run | humLN00trdo | reinaENoldest | deudescsp12075 | coFSnR2S | CoLesIRnITst | |
| French | umne | hummingbird | reina | dcu | colesir | 1st vs 5th |
| MAP | 47.57% | 45.43% | 44.58% | 41.08% | 39.51% | 20.40% |
| GMAP | 15.02% | 14.90% | 14.32% | 12.00% | 11.91% | 26.11% |
| Run | UniNEfr1 | humFR00trdo | reinaFRoldest | deudescfr12075 | CoLesIRnITst | |
| German | hummingbird | colesir | daedalus | | | 1st vs 3rd |
| MAP | 48.30% | 37.21% | 34.0% | | | 41.81% |
| GMAP | 22.53% | 14.80% | 10.61% | | | 112.35% |
| Run | humDF00trdo | CoLesIRnITst | deFSnR2S | | | |
| Italian | hummingbird | reina | dcu | daedalus | colesir | 1st vs 5th |
| MAP | 41.94% | 38.45% | 37.73% | 35.11% | 32.23% | 30.13% |
| GMAP | 11.47% | 10.55% | 9.19% | 10.30% | 8.23% | 39.37% |
| Run | humIT00trdo | reinaIToldest | deudescit1005 | nlFSnR2S | CoLesIRnITst | |
| Spanish | hummingbird | reina | dcu | daedalus | colesir | 1st vs 5th |
| MAP | 45.66% | 44.01% | 42.14% | 40.40% | 40.17% | 13.67% |
| GMAP | 23.61% | 22.65% | 21.32% | 19.64% | 18.84% | 25.32% |
| Run | humES00trdo | reinaESoldest | deudescsp12075 | coFSnR2S | CoLesIRnITst | |

Classifiche (4)

Table 9. Best entries for the robust bilingual task.

| Track | Participant Rank | | | | | Diff. |
|----------------|------------------|-------------------|--------------|-----|-----|------------|
| | 1st | 2nd | 3rd | 4th | 5th | |
| Dutch | daedalus | | | | | |
| MAP | 35.37% | | | | | |
| GMAP | 9.75% | | | | | |
| Run | nlFSnRLfr2S | | | | | |
| German | daedalus | colesir | | | | 1st vs 2nd |
| MAP | 29.16% | 25.24% | | | | 15.53% |
| GMAP | 5.18% | 4.31% | | | | 20.19% |
| Run | deFSdeRSen2S | CoLesIRendeTst | | | | |
| Spanish | reina | dcu | daedalus | | | 1st vs 3rd |
| MAP | 36.93% | 33.22% | 26.89% | | | 37.34% |
| GMAP | 13.42% | 10.44% | 6.19% | | | 116.80% |
| Run | reinaT2ESdctest | deuitqdescsp12075 | esFSesRLit2S | | | |

Classifiche (5)

Table 10. Best entries for the robust multilingual task.

| Track | Participant Rank | | | | | Diff. |
|---------------------|------------------|------------|----------------|-----------------|-----|------------|
| | 1st | 2nd | 3rd | 4th | 5th | |
| Multilingual | jaen | daedalus | colesir | reina | | 1st vs 4th |
| MAP | 27.85% | 22.67% | 22.63% | 19.96% | | 39.53% |
| GMAP | 15.69% | 11.04% | 11.24% | 13.25% | | 18.42% |
| Run | njamrsv2 | mlRSFSen2S | CoLesIRmultTst | reinaES2mtctest | | |

Hummingbird Core Tech. (Canada)

- <http://www.hummingbird.com/>
 - SearchServer
 - SearchSQL
 - Intuitive Searching
- **Ha partecipato**
 - **Monolingua**
 - Bulgaro, Francese, Ungherese, Portoghese, Inglese
 - **Robust**
 - Olandese, Inglese, Francese, Tedesco, Italiano, Spagnolo

Introduzione

- **Insieme di strumenti per lo sviluppo di applicazioni**
 - Ricerca di dati
 - Recupero di dati
- **Lavora internamente con Unicode**
- **Supporta la maggior parte**
 - Character sets
 - Linguaggi**più importanti al mondo**

Contenuti

- **Indicizzazione**
 - **cTREC**
- **Ricerca**
 - **SearchSQL**
- **Esperimenti**
 - 5 esperimenti effettuati nel Maggio 2006
 - Primary Measures Recall
 - Secondary Measures Recall
 - Risultati Statistici

cTREC (1)

- **Indicizzazione Full-Text**
- **Analizzatore dei testi**
 - Estensione rispetto alla versione 2005
- **Estrazione**
 - **DOCNO** identificatore del documento
 - Informazioni dall'header di ogni file .GOV della collezione
 - Salvataggio delle informazioni nelle colonne dalla 129 alla 140 della SearchServer Table

cTREC (2)

- q **129: "non-empty" title**
 - q Contiene il valore della prima colonna non vuota dalla 130 alla 135
- q **130: title**
 - q Se c'è, il testo è contenuto tra i tags <TITLE> e </TITLE>
- q **131: meta title**
 - q Se c'è, contiene il tag META TITLE

cTREC (3)

- q **132: meta subject**
 - q Se c'è, contiene il tag META SUBJECT
- q **133: meta description**
 - q Se c'è, contiene il tag META DESCRIPTION
- q **134: first heading**
 - q Se c'è, il testo è contenuto tra i tags <H1>, <H2>, <H3> e </H1></H2></H3>
- q **135: URL**
 - q Troncata a 256 bytes mentre le altre proprietà a 1024 bytes

cTREC (4)

- q **136: URL type, s=#slash (http:// escluso)**
 - q Homepage-type se termina con
 - q "/", "/index.html", "/index.htm"
 - q "/default.html", "/default.htm", "/default.asp"
 - q "/home.html", "/home.htm"
 - q "/welcome.html", "/welcome.htm"
 - q Tipi
 - q ROOT, s=0 e s=1
 - q SUBROOT, s=2
 - q PATH, s≥3
 - q FILE, tutte le altre URLs

cTREC (5)

- q **137: URL depth, d = s - p**
 - q n, numero dei punti prima del primo slash dopo "http://"
 - q q, numero di "?", ";", "#"
 - q p = n + q
 - q Se Homepage-type allora d-1
 - q Tipi
 - q URLDEPTH A, d=1
 - q URLDEPTH AB, d=2
 - q URLDEPTH ABC, d=3
 - q ...
 - q Fino a d=25 come limite massimo

cTREC (6)

Table 1: Examples of URL Type and Depth Values

| URL | Type | Depth | Depth Term |
|----------------------------|---------|-------|-------------|
| http://nasa.gov/ | ROOT | 1 | URLDEPTHA |
| http://www.nasa.gov/ | ROOT | 1 | URLDEPTHA |
| http://jpl.nasa.gov/ | ROOT | 2 | URLDEPTHAB |
| http://fred.jpl.nasa.gov/ | ROOT | 3 | URLDEPTHABC |
| http://nasa.gov/jpl/ | SUBROOT | 2 | URLDEPTHAB |
| http://nasa.gov/jpl/fred/ | PATH | 3 | URLDEPTHABC |
| http://nasa.gov/index.html | ROOT | 1 | URLDEPTHA |
| http://nasa.gov/fred.html | FILE | 2 | URLDEPTHAB |

cTREC (7)

138: meta keywords

- Se c'è, contiene il tag META KEYWORDS

139: tutte le proprietà (keyword escluse)

- Concatenazione delle colonne dalla 130 alla 135

140: tutte le proprietà

- Concatenazione delle colonne 139 e 138

cTREC (8)

```
create schema GOV
create table GOV
(
  DOCNO varchar(256) 128,
  NONEMPTY_TITLE varchar(2048) 129,
  TITLE varchar(2048) 130,
  META_TITLE varchar(2048) 131,
  META_SUBJECT varchar(2048) 132,
  META_DESCRIPTION varchar(2048) 133,
  FIRST_HEADING varchar(2048) 134,
  URL varchar(2048) 135,
  URL_TYPE varchar(2048) 136,
  URL_DEPTH varchar(2048) 137,
  META_KEYWORDS varchar(2048) 138,
  ALL_BUT_KEYWORDS varchar(2048) 139,
  ALL_PROPS varchar(2048) 140
)
periodic
stopfile 'mygov.stp'
basepath 'e:\data\compressed';
```

ServerSearch Table chiamata GOV
costruita per la collezione .GOV,
con lo statement SearchQuery qui
a fianco

cTREC (9)

```
insert into GOV ( ft_sfname, ft_flist )
values ( 'gov', 'cTREC/E/d=128/H:s!cTREC/w/q/@:q');
```

GOV inizialmente è vuota, viene inserita solo una riga contenente la
directory principale del data set

```
validate index GOV validate table;
```

Per indicizzare la tabella GOV usando un Validate Index

Estensione cTREC 2006

- Conversione in UNICODE
 - Lettura del charset della pagina per la conversione in UNICODE
- Apostrofo come termine separatore
- Accenti e Stop Words non vengono indicizzati
 - In Inglese ad esempio "the", "by" e "of"

Matching Supportati

- Matching Esatto
 - Tecniche di normalizzazione basate su UNICODE
 - Decomposizione
 - Conversione in maiuscolo
 - ...
- Matching Morfologico
 - Inflessioni
 - Derivazioni
 - Composizioni

→ Dipendono dal componente linguistico usato

SearchSQL (1)

- Sperimentazione predicato CONTAINS
- SearchSQL per eseguire OR booleano tra parole della query

```
SELECT RELEVANCE('2:3') AS REL, DOCNO
FROM CLEFOGEN
WHERE FT_TEXT CONTAINS 'Swiss' OR 'referendums'
ORDER BY REL DESC;
```

Diagramma di annotazione della query:

- tabella: CLEFOGEN
- rilevanza: RELEVANCE('2:3')
- identificatore documento: DOCNO
- predicato: CONTAINS
- OR: OR

SearchSQL (2)

- SearchServer nel calcolo della rilevanza
 - Riduce
 - Frequenza di termini
 - Frequenza inversa dei documenti usando un'approssimazione logaritmica
 - Regolando
 - Lunghezza del documento
 - Basati sul tema delle parole
- RELEVANCE_METHOD = 2:3
- RELEVANCE_DLEN_IMP = 750

Esperimenti (1)

- Identificatore esperimento della forma
 - Es. "humBG06tde", con
 - t, indica il campo title
 - d, indica il campo description
 - n, indica il campo narrative
 - e, indica quale query expansion è stata usata sulle tre righe usando il blind feedback (vd. dalla slide 55)
- Per molti linguaggi, usando Description e Narrative, sono rimosse parole come
 - "find", "relevant" e "document"

Esperimenti (2)

- "t", solo il campo title del topic usato
- "te", lo stesso di "t"
 - Uso del blind feedback sulle 3 righe della query "t" per espandere la query
- "td", campo title e description del topic usato
- "tde", lo stesso di "td"
 - Uso del blind feedback sulle 3 righe della query "t" per espandere la query
- "tdn", campo title, description e narrative del topic usato

Blind Feedback (1)

- Utente effettua una query
- Un insieme di documenti viene ritrovato
 - Assumiamo che siano rilevanti
 - Le parole in questi documenti sono usate per espandere la query
- Esecuzione delle query espansa
- Calcolo di un nuovo insieme di documenti da fornire all'utente

Blind Feedback (2)

- Rocchio

$$Q_{i+1} = \alpha Q_i + \beta \sum_{j=1}^{|R|} \frac{R_j}{|R_j|}$$

Qi query originale
Qi+1 query espansa
Rj insieme documenti recuperati
R documenti collezione
 α, β , segreti

- Ide Dec-Hi

$$Q_{i+1} = Q_i + \sum_{j=1}^{|R|} R_j$$

Qi query originale
Qi+1 query espansa
R documenti collezione
Rj insieme documenti recuperati

Misure di Valutazione

- In molti casi le misurazioni tradizionali non sono veritiere
 - Recall, Precision...
- Nel caso di un solo documento rilevante trovato
 - Recall, Precision raddoppiano erroneamente i loro valori
- Uso di nuove misure

Primary Measures Recall (1)

- Valori calcolati sul primo documento rilevante di un topic
- **Generalized Success@30 (GS30)**, per un topic è 1.024^{1-r}
- **Generalized Success@10 (GS10)**, per un topic è 1.08^{1-r}
- r rank della riga per cui una pagina richiesta è stata trovata, 0 nel caso una pagina richiesta non sia stata trovata

Primary Measures Recall (2)

- **Success@n (S@n)**, per un topic è
 - 1 se la pagina richiesta si trova tra le prime n ottenute 0 altrimenti
 - $n=1$ e $n=10$
- **Reciprocal Rank (RR)**, per un topic è $1/r$
 - **MRR**, media dei reciprocal rank $RR/\#topics$
- r rank della riga per cui una pagina richiesta è stata trovata, 0 nel caso una pagina richiesta non sia stata trovata

Secondary Measures Recall

- Valori calcolati sugli altri documenti rilevanti di un topic
- **Precision@n**, precisione dopo aver trovato n documenti ($n=10$)
- **Average Precision AP**, Geometric MAP
- **GMAP'**, GMAP linearizzata che mappa i valori nell'intervallo $[0,1]$

Risultati Statistici

▫ Esempio di tabella riassuntiva

- Valutazioni delle primary e secondary measures recall per la lingua bulgara

Table 3: Mean Scores of Monolingual Ad Hoc Runs

| Run | GS30 | GS10 | S10 | MRR | S1 | P10 | GMAP | MAP |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|
| (humBG06tdn) | 0.919 | 0.846 | 46/50 | 0.651 | 26/50 | 0.322 | 0.200 | 0.305 |
| humBG06td | 0.903 | 0.829 | 44/50 | 0.648 | 26/50 | 0.308 | 0.172 | 0.285 |
| humBG06tde | 0.896 | 0.826 | 44/50 | 0.616 | 23/50 | 0.334 | 0.182 | 0.305 |
| (humBG06te) | 0.858 | 0.742 | 38/50 | 0.537 | 21/50 | 0.314 | 0.148 | 0.291 |
| humBG06t | 0.832 | 0.724 | 38/50 | 0.513 | 19/50 | 0.282 | 0.108 | 0.261 |

Primary Measures Recall Secondary Measures Recall

Risultati Query Expansion (1)

▫ Espansione "Title" Queries

- Description
- Blind Feedback

▫ Espansione "Title+Desc" Queries

- Narrative
- Blind Feedback

Risultati Query Expansion (2)

▫ Espansione "Title" Queries

- Description
 - Incremento per GS30, GS10, MRR, P10, GMAP e MAP
 - Tecnica robusta in alcuni casi migliora gli scarsi risultati dell'uso del solo Title
- Blind Feedback
 - Nessun incremento per GS30, GS10 e MRR
 - Incremento per P10, GMAP e MAP
 - Tecnica non robusta perché è poco probabile migliorare i risultati scarsi

Risultati Query Expansion (3)

▫ Espansione "Title+Desc" Queries

- Narrative
 - Incremento per GS30, GS10, MRR, P10, GMAP e MAP
 - Non tanto quanto la coppia "Title+Desc"
- Blind Feedback
 - Incremento per P10, GMAP e MAP
 - Incremento per GS30, GS10, MRR
 - Ma solo per l'Ungherese
 - Possibili errori

Risultati Query Expansion (4)

- **Espansione "Title" Queries**
 - **Description vs. Blind Feedback**
 - Ogni Primary Measures Recall ha almeno un cambiamento statistico positivo
 - Description mantiene robustezza, Blind Feedback no
 - Nessun significativo cambiamento Secondary Measures Recall
- **Espansione "Title+Desc" Queries**
 - **Narrive vs. Blind Feedback**
 - Narrative modifiche modeste alle Primary Measures Recall
 - Blind Feedback modifiche maggiori alle Secondary Measures Recall

Risultati Robust Task (1)

- **Per ogni lingua**
 - Una prova per "td" senza blind feedback
 - Una prova per "tde" con blind feedback
 - GMAP ha valore più alto rispetto a tutti gli altri casi
- **Blind Feedback è risultato**
 - Negativo per le Primary Measures Recall
 - Positivo con notevoli incrementi per le Secondary Measures Recall
 - GMAP parametro pubblicizzato dal CLEF

Risultati Robust Task (2)

Table 10: Mean Scores of Robust "Training" and "Test" Runs

| Run | GS30 | GS10 | S10 | MRR | S1 | P10 | GMAP | MAP |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|
| lumDE06Rtd0 | 0.963 | 0.923 | 59/60 | 0.747 | 37/60 | 0.530 | 0.322 | 0.422 |
| lumDE06Rtde0 | 0.966 | 0.932 | 59/60 | 0.786 | 41/60 | 0.553 | 0.367 | 0.470 |
| lumEN06Rtd0 | 0.887 | 0.795 | 46/55 | 0.660 | 32/55 | 0.342 | 0.204 | 0.385 |
| lumEN06Rtde0 | 0.887 | 0.789 | 45/55 | 0.666 | 33/55 | 0.378 | 0.225 | 0.424 |
| lumES06Rtd0 | 0.947 | 0.896 | 55/59 | 0.761 | 40/59 | 0.464 | 0.263 | 0.367 |
| lumES06Rtde0 | 0.924 | 0.871 | 53/59 | 0.729 | 37/59 | 0.485 | 0.285 | 0.407 |
| lumFR06Rtd0 | 0.908 | 0.858 | 54/60 | 0.702 | 35/60 | 0.337 | 0.241 | 0.403 |
| lumFR06Rtde0 | 0.912 | 0.856 | 54/60 | 0.682 | 33/60 | 0.357 | 0.269 | 0.432 |
| lumIT06Rtd0 | 0.915 | 0.825 | 53/57 | 0.571 | 23/57 | 0.360 | 0.209 | 0.357 |
| lumIT06Rtde0 | 0.910 | 0.817 | 53/57 | 0.556 | 22/57 | 0.389 | 0.232 | 0.392 |
| lumNL06Rtd0 | 0.959 | 0.902 | 56/60 | 0.776 | 42/60 | 0.492 | 0.342 | 0.438 |
| lumNL06Rtde0 | 0.953 | 0.901 | 57/60 | 0.757 | 39/60 | 0.503 | 0.373 | 0.475 |
| lumDE06Rtd | 0.964 | 0.904 | 91/95 | 0.706 | 54/95 | 0.483 | 0.330 | 0.447 |
| lumDE06Rtde | 0.953 | 0.891 | 91/95 | 0.711 | 56/95 | 0.539 | 0.382 | 0.508 |
| lumEN06Rtd | 0.952 | 0.899 | 82/88 | 0.737 | 54/88 | 0.382 | 0.373 | 0.486 |
| lumEN06Rtde | 0.952 | 0.900 | 82/88 | 0.760 | 58/88 | 0.435 | 0.419 | 0.541 |
| lumES06Rtd | 0.945 | 0.879 | 90/97 | 0.702 | 56/97 | 0.473 | 0.302 | 0.433 |
| lumES06Rtde | 0.917 | 0.850 | 86/97 | 0.706 | 50/97 | 0.500 | 0.322 | 0.471 |
| lumFR06Rtd | 0.944 | 0.888 | 86/91 | 0.709 | 53/91 | 0.412 | 0.317 | 0.470 |
| lumFR06Rtde | 0.951 | 0.895 | 85/91 | 0.732 | 57/91 | 0.430 | 0.386 | 0.499 |
| lumIT06Rtd | 0.940 | 0.885 | 84/90 | 0.676 | 46/90 | 0.407 | 0.281 | 0.409 |
| lumIT06Rtde | 0.945 | 0.892 | 85/90 | 0.692 | 48/90 | 0.447 | 0.324 | 0.466 |
| lumNL06Rtd | 0.954 | 0.907 | 92/96 | 0.785 | 68/96 | 0.478 | 0.346 | 0.484 |
| lumNL06Rtde | 0.948 | 0.899 | 89/96 | 0.779 | 67/96 | 0.531 | 0.393 | 0.532 |

Conclusioni

- **GS30, GS10, S@n e MRR**
 - Riflettono la robustezza dei sistemi
- **Blind Feedback non funziona con la robustezza**
 - Usati per il calcolo delle Secondary Measure Recall
 - GMAP, MAP e P10
- **Risultati pratici a cui stanno seguendo risultati teorici che rafforzano queste conoscenze**

Sitografia

- ¶ <http://www.clef-campaign.org/>
- ¶ http://www.clef-campaign.org/2006/working_notes/CLEF2006WN-Contents.html
- ¶ <http://roquefort.di.unipi.it/%7Eferrax/Teach/InformationRetrieval/QualitySE.pdf>
- ¶ <http://roquefort.di.unipi.it/%7Eferrax/Teach/InformationRetrieval/TextualRanking.pdf>
- ¶ http://trec.nist.gov/pubs/trec14/appendices/CE_MEASURES05.pdf
- ¶ http://en.wikipedia.org/wiki/Information_retrieval#Average_precision
- ¶ <http://www.csee.umbc.edu/~ian/irF02/lectures/09Evaluation.pdf>
- ¶ http://www.clef-campaign.org/2006/working_notes/workingnotes2006/tomlinsonAd-HocCLEF2006.pdf
- ¶ <http://trec.nist.gov/pubs/trec11/papers/hummingbird.tomlinson.pdf>
- ¶ http://www.clef-campaign.org/2005/working_notes/workingnotes2005/tomlinsonadhoc05.pdf
- ¶ <http://filebox.vt.edu/users/wfan/paper/ARRANGER/p52-Fan.pdf>

Bibliografia

- ¶ "CLEF 2006: Ad Hoc Track Overview" - G. M. Di Nunzio, N. Ferro, T. Mandl and C. Peters
- ¶ "Evaluation of a search engine: Precision vs. Recall, Interpolated precision and F-measure" - P. Ferragina
- ¶ "Textual document scoring and fast cosine calculation, Relevance feedback and Rocchio's method" - P. Ferragina
- ¶ "Common Evaluation Measures" - TREC 2005
- ¶ "IR Evaluation (Lecture 9)" - Dr. I. Soboroff
- ¶ "Comparing the Robustness of Expansion Techniques and Retrieval Measures" - S. Tomlinson
- ¶ "European Web Retrieval Experiments with Hummingbird SearchServerTM at CLEF 2005" - S. Tomlinson
- ¶ "Experiments in Named Page Finding and Arabic Retrieval with Hummingbird SearchServerTM at TREC 2002" - S. Tomlinson
- ¶ "Tuning Before Feedback: Combining Ranking Discovery and Blind Feedback for Robust Retrieval" - W. Fan, M. Luo, L. Wang, W. Xi, E. A. Fox