

Experiments with PROSIT
at the Fondazione Ugo Bordoni

Gianni Amati, Claudio Carpineto, and Giovanni Romano

Fondazione Ugo Bordoni

Roma

carpinet@fub.it

Overview

- **PROSIT's model**
- **Earlier implementations of PROSIT**
- **PROSIT at CLEF02: adding proximity and coordination level-based matching**
- **Results**
- **Conclusions**

$$W = \text{Inf}_1 \cdot \text{Inf}_2$$



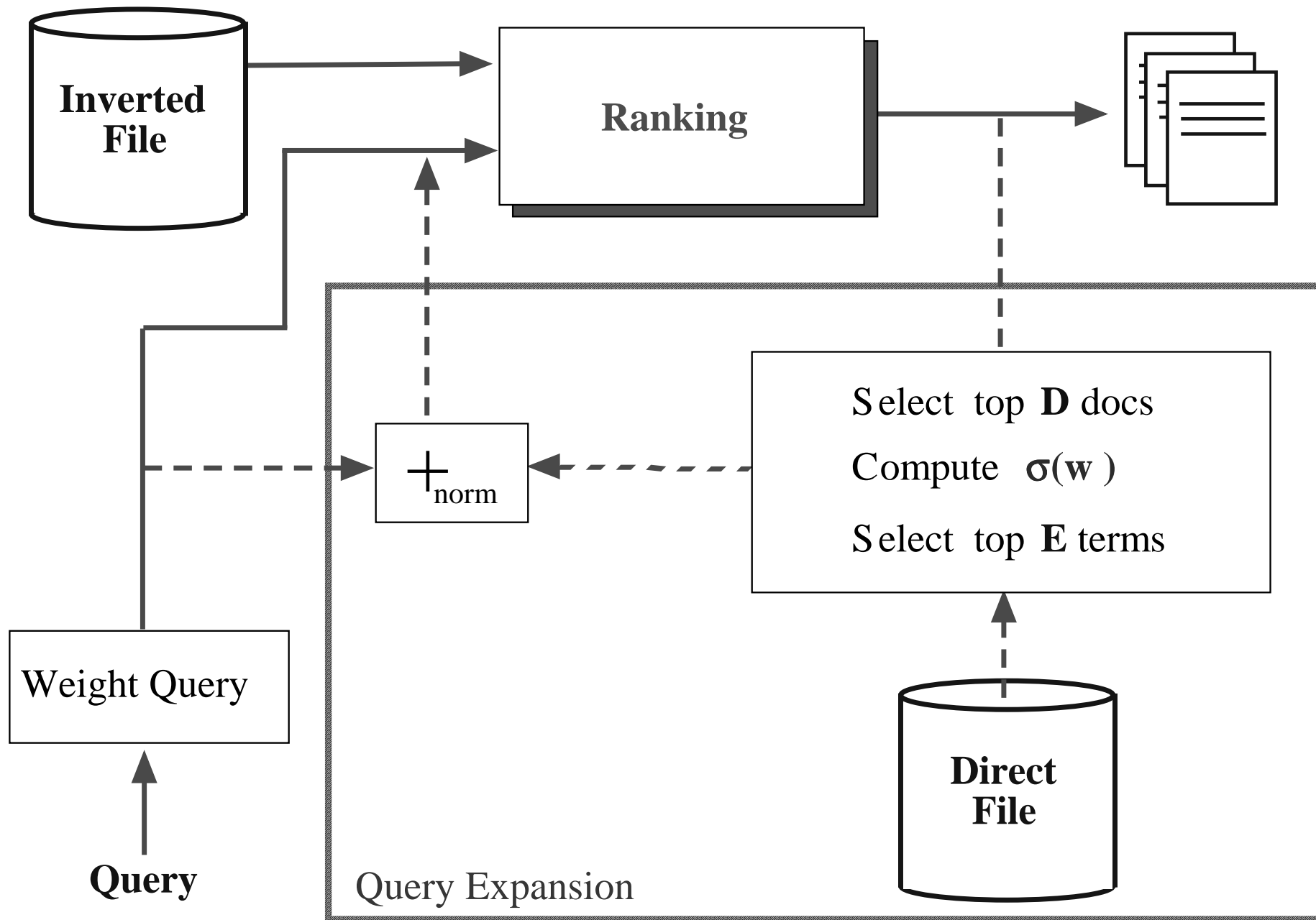
$$tf \cdot \log [(N + 1) / (n + 0.5)]$$

.....

$$1/(tf + 1)$$

.....

$$tf_n = tf \cdot \log (1 + K \cdot avg_l / l)$$



Earlier implementations of PROSIT's IR model

- **Best system at TREC-10 Web Track for the topic relevance task**
- **Search engine of the web site of the Italian Ministry of Communications (www.comunicazioni.it)**



licenza radioamatore

Cerca nel sito

Powered by [Fondazione Ugo Bordoni](#)

Trovati 66 Documenti

- 1 [| URP | UTILITA' PER IL CITTADINO E LE IMPRESE | Per diventare Radioamatore |](#)
www.comunicazioni.it/it/index.php?Mn1=6&Mn2=28&Mn3=18
- 2 [| URP | UTILITA' PER IL CITTADINO E PER LE IMPRESE | Per diventare C.B. |](#)
www.comunicazioni.it/it/index.php?Mn1=6&Mn2=28&Mn3=19
- 3 [| URP | UTILITA' PER IL CITTADINO E PER LE IMPRESE | Per ottenere titoli di abilitazione |](#)
www.comunicazioni.it/it/index.php?Mn1=6&Mn2=28&Mn3=20
- 4 [| URP | UTILITA' PER IL CITTADINO E PER LE IMPRESE | Servizio Audiotex |](#)
www.comunicazioni.it/it/index.php?Mn1=6
- 5 [Regolamento recante norme per il recepimento](#)
www.comunicazioni.it/it/index.php?IdPag=341
- 6 [| URP | Per ottenere licenze individuali per i servizi postali |](#)
www.comunicazioni.it/it/index.php?Mn1=6&Mn2=28&Mn3=26
- 7 [| URP | UTILITA' PER IL CITTADINO E PER LE IMPRESE | Per diventare operatori per TLC |](#)
www.comunicazioni.it/it/index.php?Mn1=6&Mn2=28&Mn3=24
- 8 [Glossario](#)
www.comunicazioni.it/it/index.php?IdPag=355
- 9 [| URP | Corsi di formazione |](#)
www.comunicazioni.it/it/index.php?Mn1=6&Mn2=28&Mn3=29
- 10 [| URP | Archivio CAP |](#)
www.comunicazioni.it/it/index.php?Mn1=6&Mn2=28&Mn3=30

1-10 [11-20](#) [21-30](#) [31-40](#) [41-50](#) [51-60](#) [61-66](#)

Adding bigrams to PROSIT

***Technique:* Lexical affinities**

***Implementation:* Generation at run time**

***Results:* Bigram scores ineffective when used alone or combined with final unigram scores, better results when used to drive query expansion**

Reranking PROSIT's retrieval results with Coordination Level-based Matching

***Technique:* Documents partially ordered according
to their CLM + ties broken with PROSIT scores**

***Results:* Partially good results only when using
full CLM**

Retrieval effectiveness of FUB runs

	PROSIT	PROSIT + bigrams	PROSIT + CLM	PROSIT + bigrams + CLM
CLEF 2001				
CLEF 2002				

Conclusions

- **PROSIT's IR model highly effective**
- **Bigrams moderately effective**
- **CLM mostly ineffective**