

# What Can Language Technology do for IR?

ECIR 2008  
Industry Day

Antonio Valderrábanos  
Bitext  
<http://www.bitext.com>  
[info@bitext.com](mailto:info@bitext.com)



# Table of contents

- Bitext
- Language Technology and Search Engines
  - DOES/COULD DO/IS DOING
- What Bitext is offering for Search Engines
  - Live + NaturalFinder
- What Bitext is offering for other Search Environments
- Research at Bitext



# Bitext - Applied Language Technologies

- Before Bitext – Since late 80's
  - IBM, WorPerfect, Novell
  - LTI - Carnegie Mellon University
  - Information Retrieval
  - Grammar checking
  - Assisted Translation
- Early Bitext - EU funded projects - IST
  - LIQUID, ALLES, TT2
- Bitext – Developer of complementary products
  - NaturalFinder + Search Engines



# What Language Technology DOES for IR

- Nothing
  - Query side
    - *"Special offers for..."*
    - 53 70 65 63 69 61 6C 20 6F 66 66 65 72 73
  - Index
    - 53 70 65 63 69 61 6C
    - 6F 66 66 65 72 73
- or almost nothing as we'll see...



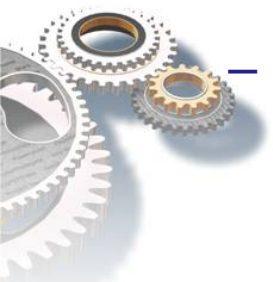
# What Language Technology COULD DO for IR

- Everything - Search is a language problem!
  - Query side
    - *“Special offers for MP3 players”*
    - *offers*, POS:noun, NUM:plural... SYN: bargain
  - Index
    - *special*, adj...
    - *offers*, noun...
    - ...NOUN PHRASE
- but it is not happening! Or it is?



# What Language Technology IS DOING for IR

- Actually, there is something
  - Spelling – Live, Google, Yahoo – Not Amazon!
    - Avoids silence
  - Stemming – Live, Google
    - Increases recall
  - Natural language interface – Hakia, Powerset
- And probably, there's more coming
- (Debate: Symbolic vs. Statistical Techniques)



# What Bitext is offering for Search Engines

- As much as linguistic technology can do
  - One integrated NLP suite: DataSuite
    - Spelling – DataSpell
    - Stemming – DataLexica
    - Semantics – DataNet
    - Morph-syntactic processing - DataGrammar
  - Packed solutions for specific applications:
    - NaturalFinder, for search engines
  - For Spanish, English, Basque, Catalan...



# NaturalFinder, add-on for any search engine

- NaturalFinder, functionality
  - Natural Language Interface
    - “Italian painters that lived in Spain”*
      - Better user experience
      - No keywordese, no two-word queries, no booleans
  - Linguistic Knowledge – Transparent for the user
    - Better results
    - Better ranking (recall & precision)





# NaturalFinder, add-on for any search engine

- NaturalFinder, technology
  - Integrates in a non-intrusive way
    - string-based interface –API
    - in any platform -Windows, Linux/UNIX
  - Already integrated with
    - Google Search Appliance
    - dtSearch
    - Memex
    - Microsoft SharePoint
    - Oracle Secure Enterprise Search...



# NaturalFinder, add-on for any search engine

- NaturalFinder + Live
  - Architecture
    - Live: Docs, Index, Search Engine
    - NaturalFinder: User Interface
  - Limitations
    - Under development
    - Query length
    - Case and accent sensitivity
    - Operator nesting – NOT
- Thanks to Microsoft!





information on tourist attractions in the uk but not in London

NF+Live Live

Examples Advantages Features

Find what you are looking for in Live using **NaturalFinder**. You will retrieve all relevant documents for your query with just one click and in your own words

Try now

- Type your query in natural language:  
**information about tropical paradises which are not in the Pacific**
- Click on '**Live**'. The application will launch the query using Live search engine.
- Now, try the same query using our linguistic technology by clicking on '**NF+Live**', and check the difference.

Other examples of queries with NaturalFinder

- **information on sharks in the Red Sea**
- **outcome of the Spanish general election in 2000**
- **information on UFOs in southern Spain**
- **terrorist groups supported by Syrian intelligence agency**

Demos de Bitext.com - Tecnologías de Lenguaje Natural - Windows Internet Explorer

http://demos.bitext.com/LIVE-EN/search.aspx?\_\_VIEWSTATE=%2FwEPDwUKMjAxNTYyNjE5NWRkc1a1Ozmf56ObiU6AoW%2FWpxOJM4%3D&maxFiles=10&sort=Hits&cmd=se

Live Search provisto por MSN

Archivo Edición Ver Favoritos Herramientas Ayuda

Buscar web...

Demos de Bitext.com - Tecnologías de Lenguaje Natural

Página Herramientas

## LIVE without NaturalFinder

### British Museum London

British Museum. Britain guide, tourist information UK ... exhibitions, although not to ... The British Museum in London UK is one of the tourist attractions in the United ...

### London Dungeon UK

London Dungeon. Britain guide, tourist information UK ... it is not recommended ... The London Dungeon in England UK is one of the tourist attractions in the United ...

### London Hotels | London Holidays | London Tourist Information | London ...

... is soaked in such a rich imperial history that is not found anywhere else. On VisitLondon.Net, you will find Hotels in London, tourist information, restaurants, events, attractions ...

### London Tourist Information one of the most popular destinations in the ...

... London Tourist Information ... tourist attractions ... not forgetting the spectacular London Eye. The West End. Theatreland. Art Galleries and Museums. The Tate. Lots to see and do. London Tourist Information

### UK-Tourist-Info.com: Tourist Attractions & Information for Places in ...

Tourist attractions and information for places in London England, Scotland and Wales. ... Contrary to popular belief, the UK does not include the islands of ...

### TOURIST NET UK - Holiday Information - Hotels Guesthouses b&Bs ...

... the UK tourist with the information they need to plan ... promote tourism in the UK, they often do not provide the information you ... London Attractions | London Heritage | ...

London Accommodation Hotels - Tourist Net UK

Inicio

Windows Live Messen... Demos de Bitext.com ... Inbox - Microsoft Out... Microsoft PowerPoint ...

Internet 100%

ES 11:06



Demos de Bitext.com - Tecnologías de Lenguaje Natural - Windows Internet Explorer

http://demos.bitext.com/LIVE-EN/search.aspx?\_\_VIEWSTATE=%2FwEPDwUKMjAxNTYyNjE5NWRkc1a1Ozmf56ObiU6AoW%2FWpxOJM4%3D&maxFiles=10&sort=Hits&cmd=se

Live Search provisto por MSN

Archivo Edición Ver Favoritos Herramientas Ayuda

Buscar web...

Demos de Bitext.com - Tecnologías de Lenguaje Natural

Página Herramientas

### LIVE with NaturalFinder

#### Historic Sites in Cornwall - Visit Cornwall

... collections of historic sites in the UK. But don't ... Tourist Attractions ... Visitor Information

#### Runcorn Cheshire information for tourists and visitors from TourUK

... the most industrialised parts of the UK, but ... Norton Priory is Runcorn's main tourist attraction. ... To See - Runcorn Tourist Attractions

#### Guernsey Hotels, Guernsey Bed & Breakfast. We have Pubs & Inns in ...

... accommodation in Guernsey and good UK Offshore tourist attractions ... It is not a part of the UK but is a separate ... Guernsey Information: Travel help Architects Surveyors

#### Cornwall Holiday Resource | Cornwall

Cornwall information on resorts towns and villages. ... not only popular with holiday makers from the UK but ... and Poldark tin mines have been reopened as a tourist attraction.

#### The Dedicated Partnership - Poster Distribution Costings

... by the company extends throughout the UK but can be adapted to meet individual customers' needs. We cover tourist information centres, tourist attractions ...

#### Salisbury District Council | Ten-year action plan to help local ...

We boast top tourist attractions such as Salisbury Cathedral ... the most beautiful countryside found in the UK. "But it ... For more information and to request a copy of the ...

#### Darwin Australia - Getting Around An Australian Outback City

Internet 100%

Inicio Windows Live Messen... Demos de Bitext.com ... Inbox - Microsoft Out... Microsoft PowerPoint ...

E5 11:05



# What Bitext is offering for Search

- Other packed solutions for search oriented environments
  - Geographic Information Systems
    - NaturalGIS
  - Customer Relationship Mgmt
    - NaturalAssistant
  - Databases
    - NaturalSQL
  - No-index - Regular expressions



# What Bitext is offering for Search

- Pre-production Research
  - Named entity recognition with attributes  
“ACME Vice-President John Doe”
  - Text categorization
    - open-closed categories
    - known-unknown categories
    - single and multiple categories
    - paragraph and document level



# What Bitext is offering for Search

- Basic-Research
  - We are looking for a partner to develop new ways of indexing that include linguistic knowledge
  - After document metadata we need Word metadata





# Thanks!

**Antonio Valderrábanos**  
**Bitext - The Bits and Text Company**  
**Phone: +34 911461660**  
**info@bitext.com**  
**<http://www.bitext.com>**

