What Can Language Technology do for IR?

ECIR 2008
Industry Day

Antonio Valderrábanos
Bitext
http://www.bitext.com
info@bitext.com

Bitext.com
THE BITS AND TEXT COMPANY

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.com
THE BITS AND TEXT COMPANY

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

Bitext.com
THE BITS AND TEXT COMPANY

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…

Bitext.com
THE BITS AND TEXT COMPANY

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 happenning! Or it is?

Bitext.com
THE BITS AND TEXT COMPANY

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)

Bitext.com
THE BITS AND TEXT COMPANY

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…

Bitext.com
THE BITS AND TEXT COMPANY

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)

Bitext.com
THE BITS AND TEXT COMPANY

NaturalFinder, add-on for any search engine

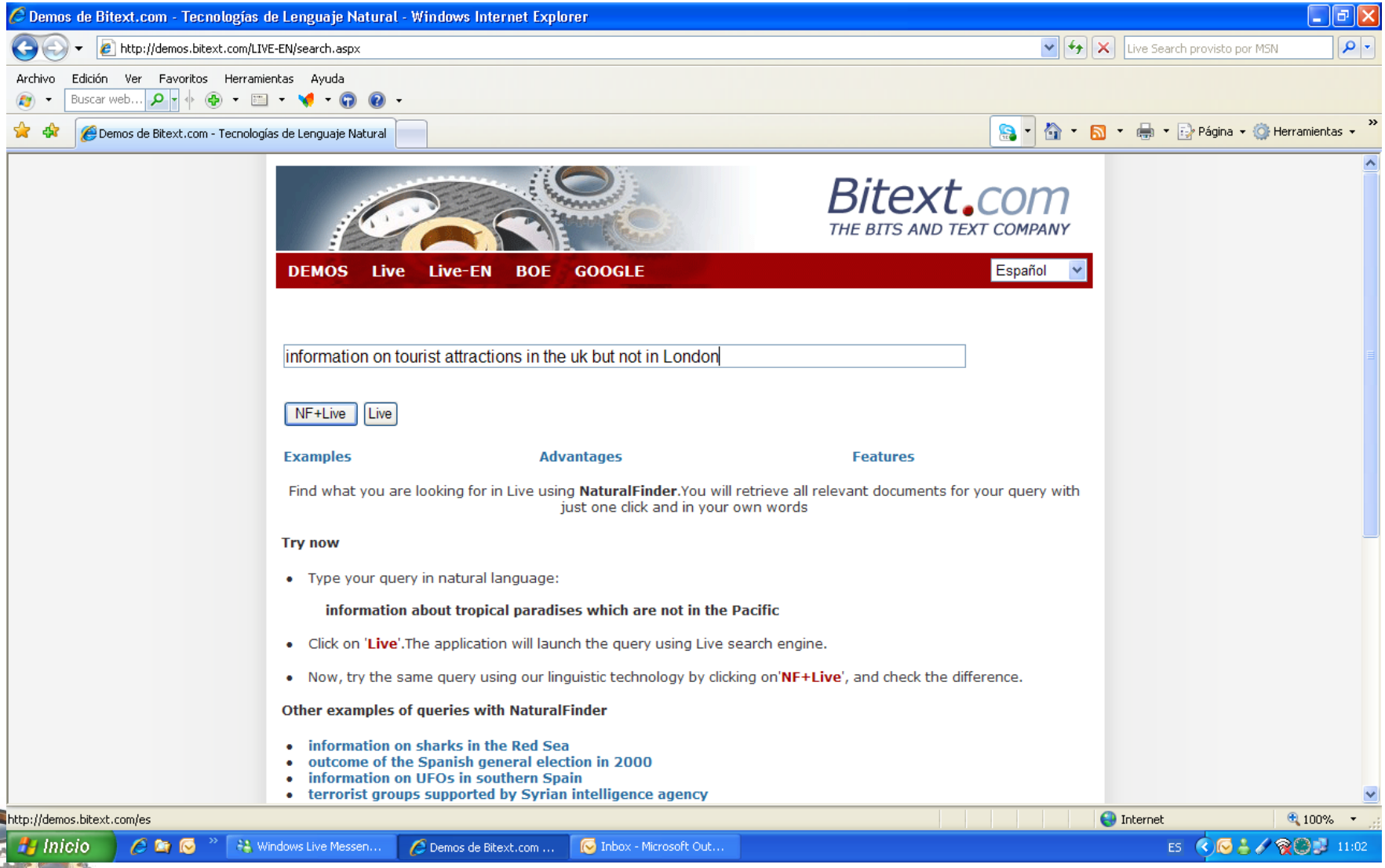
- NaturalFinder, technology
  - Integrates in a non-intrusive way
    - string-based interface –API
    - in any platform -Windows, Linux/UNIX
  - Already integrated with
    - Google Search Appliance
    - dtSearch
    - Memex
    - Microsoft SharePoint
    - Oracle Secure Enterpise Search…

Bitext.com
THE BITS AND TEXT COMPANY

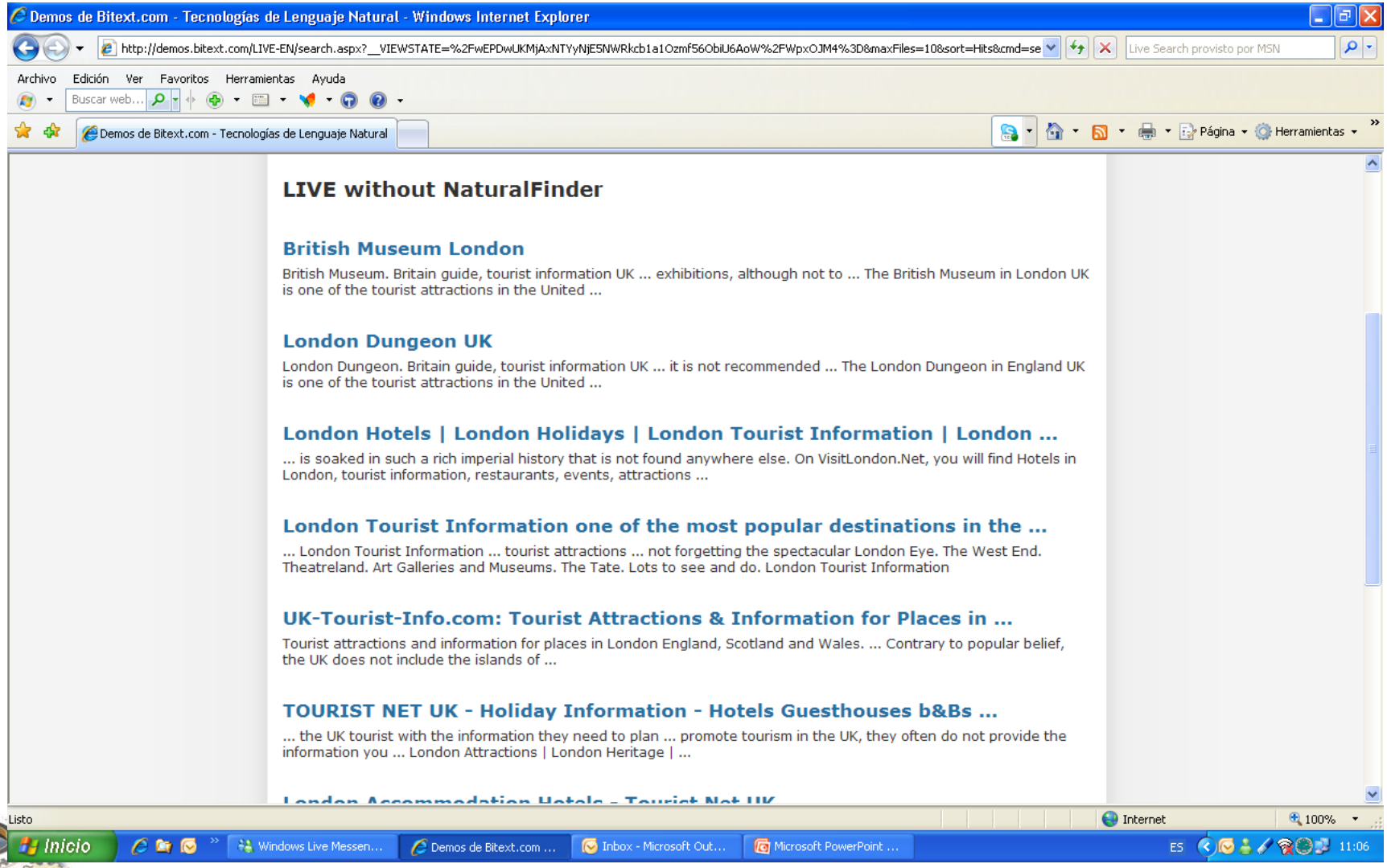
NaturalFinder, add-on for any search engine

- 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!

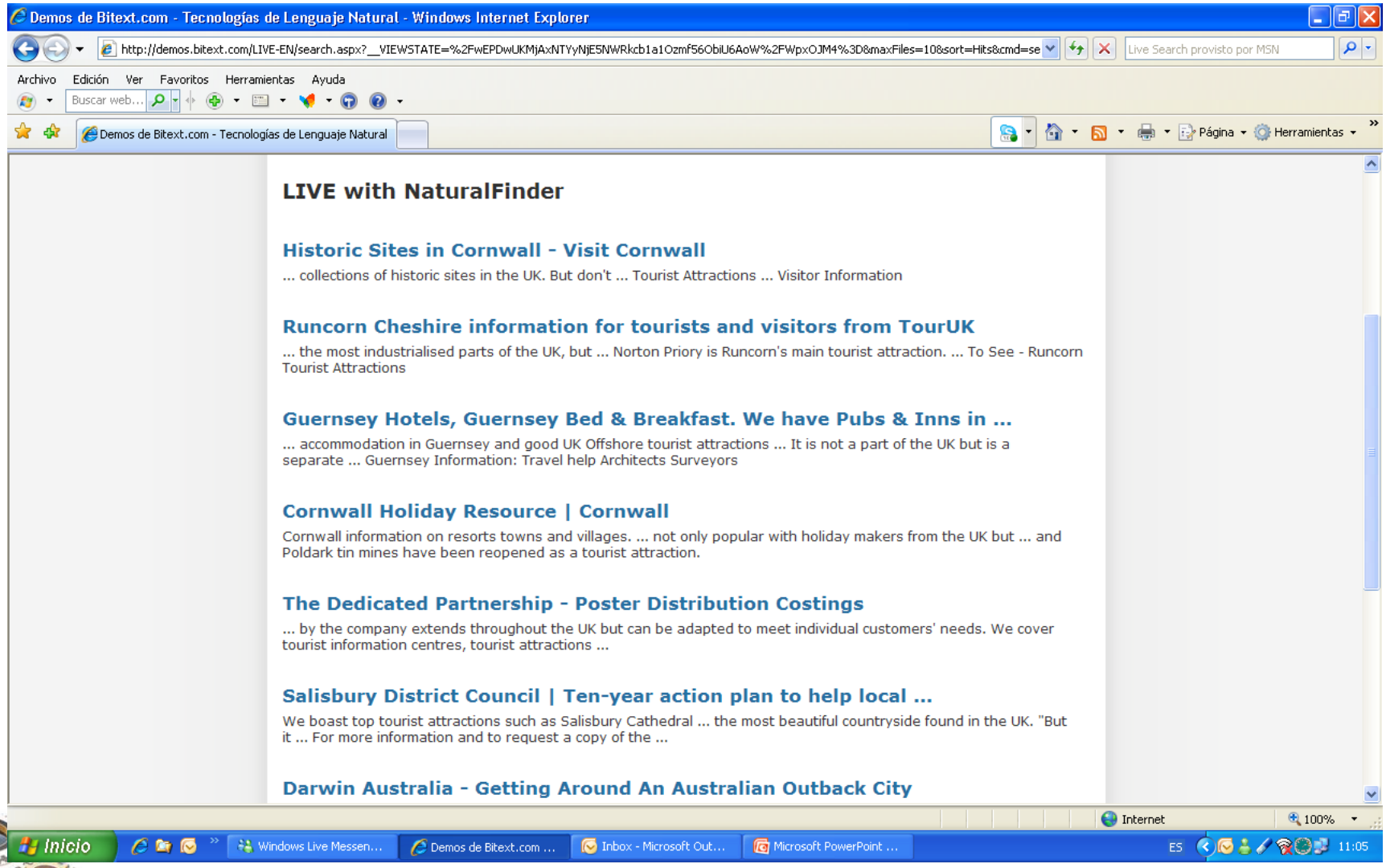
Bitext.com
THE BITS AND TEXT COMPANY



Bitext.com
THE BITS AND TEXT COMPANY



Bitext.com
THE BITS AND TEXT COMPANY



Bitext.com
THE BITS AND TEXT COMPANY

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

Bitext.com
THE BITS AND TEXT COMPANY

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

Bitext.com
THE BITS AND TEXT COMPANY

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

Bitext.com
THE BITS AND TEXT COMPANY

Thanks!

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

Bitext.com
THE BITS AND TEXT COMPANY