

# Bitext - The Bits and Text Company

## NaturalFinder - Language Technology and Search



**Antonio Valderrábanos**  
Founder & CEO  
[www.bitext.com](http://www.bitext.com)  
[info@bitext.com](mailto:info@bitext.com)

**Bitext.com**  
THE BITS AND TEXT COMPANY

---

# Table of contents

- Search market - Intranets and websites
- Natural Language Technologies in the search market
- A bit about Bitext
- NaturalFinder - Language-awareness for any search engine
  - Advantages and functionality
- Integration
  - Architecture
  - Technology
- Proposal



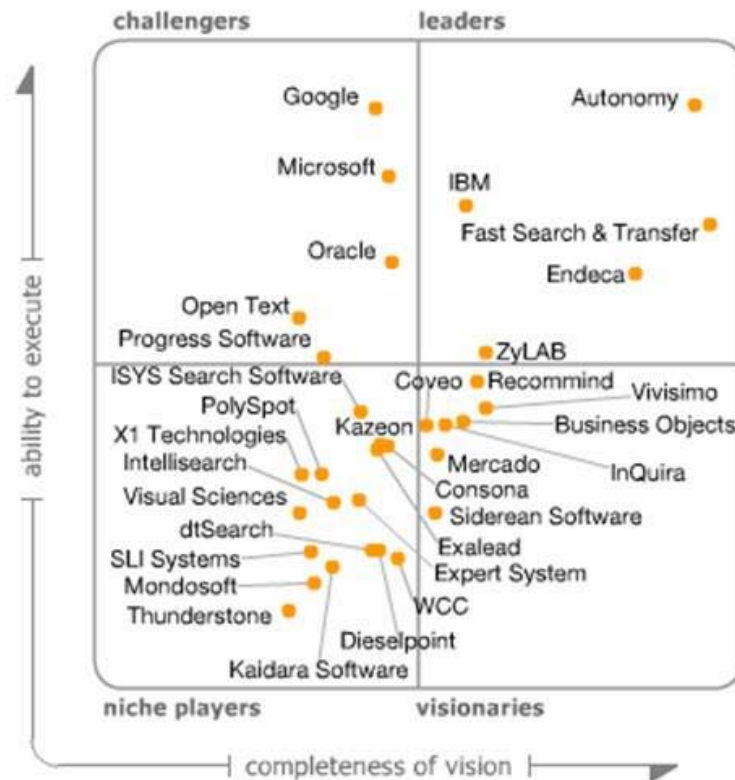
# Search Market

- ...is large, profitable and growing
- Market for enterprise unstructured data management
  - IDC - \$5.4 billion and growing at 30%
  - OVUM - \$8.8 billion and growing at 33%.
  - IDC - top 500 companies spend \$31 billion to search unstructured data
- These figures include the market for enterprise search, but do not include other search-related sectors



# Search Market

- ...is crowded: +30 players - Gartner Information Access - Sept. 2007



As of August 2007



# Search Market

- ...with room for improvement: frustrated users
- AIIM Market IQ on Findability - May 2008
  - ... 49% - it is a difficult and time consuming process to find the information they need to do their job
- Gilbane Report - April 2008
  - "What to do when your search engine doesn't work"
- "The State of Search" by Autobytel Inc., October 2007
  - 20% - inability to understand keywords
  - 3rd reason (after information overflow and paid links)



# Natural Language Technologies in the search market

- Natural Language is part of the offer of Information Access
  - Companies that integrate own technology - Gartner 2007
    - Consona
    - InQuira
  - Companies that use third-party NL technologies
    - IBM & iPhrase
    - Business Objects & Inxight
    - SPSS & Teragram
    - Powerset & Inxight & Microsoft
    - Reuters & ClearForest
  - Companies that use it for some time now
    - Autonomy
    - Google

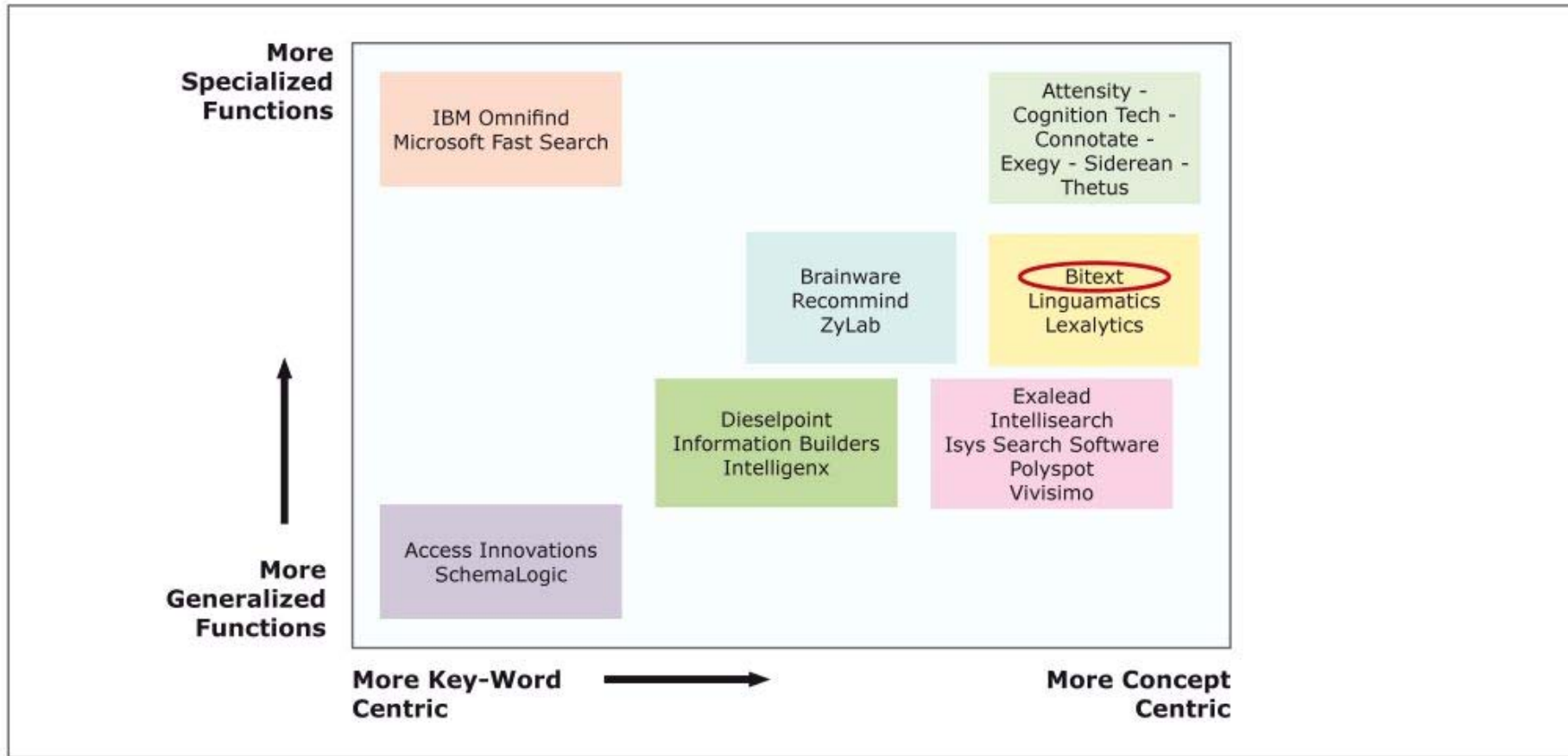


## A bit about Bitext - First steps in the US market

- Agreement with dtSearch
- Agreement with Google Search Appliance integrator in Europe
- Integration of NaturalFinder + Oracle SES – Ongoing agreement
- Gilbane Report
  - Beyond Search: What to do When Your Enterprise Search System Doesn't Work
- Stephen Arnold's blog
  - Search Wizards Speak
- Partners in Spain opening office in US



# A bit about Bitext - First steps in the US market



**Figure 20: "Beyond Search" Market Sector Vendors.**

©2008 Gilbane Group, Inc.    Página 90    <http://www.gilbane.com>

## A bit about Bitext - History

- Group of experts in linguistics with international experience
- R&D Departments
  - IBM, WordPerfect, Novell
  - LTI-Carnegie Mellon University
- European R&D Projects: European Commission IST Program
  - LIQUID, first multilingual search engine in Europe
  - ALLES, computer-assisted language learning
  - TT2, assisted and machine translation
  - EuroWordNet, semantic networks
- In 2008 the bank Caja Navarra becomes financing partner



## A bit about Bitext - Solutions Overview

- Bitext provides NL technologies to any third-party application with a two-side solution
  - Natural Language Interfaces
    - Search engines - NaturalFinder
    - Databases
    - Single-box interfaces
    - GIS
  - Text Analytics - Knowledge Extraction
    - Named entities
    - Concepts



## Solutions Overview

- Our goal
  - Provide language-awareness to any third-party software application
- Emphasizing
  - Easy integration
  - Quick development of new functionality
- One application, DataSuite
  - Independent of the task: full and chunk parsing
  - Independent of the application: search engines, GIS...
  - Independent of the language: English, Spanish...
  - Independent of the platform



## NaturalFinder – Advantages summary

- At query level
  - Minimizing user effort and maximizing user support
- At result level
  - Best results and ranking (precision)
- In short
  - NaturalFinder opens the use of advanced search options to any user



## NaturalFinder – Advantages

- **Adapts search engine to user's language, not vice versa!**
  - Natural language interface
- Easier and more effective for the user
  - information on Paris but not Hilton
  - mp3 or mp4 players
- More informative queries for the search engine!



## NaturalFinder – Advantages

- **Improved result ranking**
  - Q - chinese restaurant New York
  - A - Chinese man killed in New York restaurant
  
  - Q - Chinese restaurant in New York
  - A - Chinese restaurants reviews – New York
- Even for two-word queries
  - restaurant Atlanta
  - restaurant in Atlanta



## NaturalFinder – Advantages

- **Reduces search iterations**
  - Words which are linguistically related
    - merged >> merging, merges, merge,... Not “merganser”
  - Semantically related concepts
    - advantages >> benefits
    - Second-hand car >> used automobile



## NaturalFinder – Advantages

- **Spell checking**
  - Avoids “No results found”
    - recieved
    - immediate
  - Linguistically-motivated alternatives
  - Improves Google's "Did you mean..."



## NaturalFinder – More Advantages

- Integration of different information sources
  - With different search engines or databases
- Integration of different information datatypes:
  - un-semi-structured
- **Single-box search for any information source - NaturalBox**
  - Several boxes for one query (drop-downs, text...)
  - Examples: [www.dothomes.com](http://www.dothomes.com)



## NaturalFinder – Offline services

- **Knowledge extraction for asset leverage**
  - Log analysis at concept level
  - Improves knowledge of user behaviour
  - Extraction of named entities, concepts... (thesaurus, ontologies...)
  - Improves search engine performance
  - Document categorization - Metadata creation
  - Improves findability



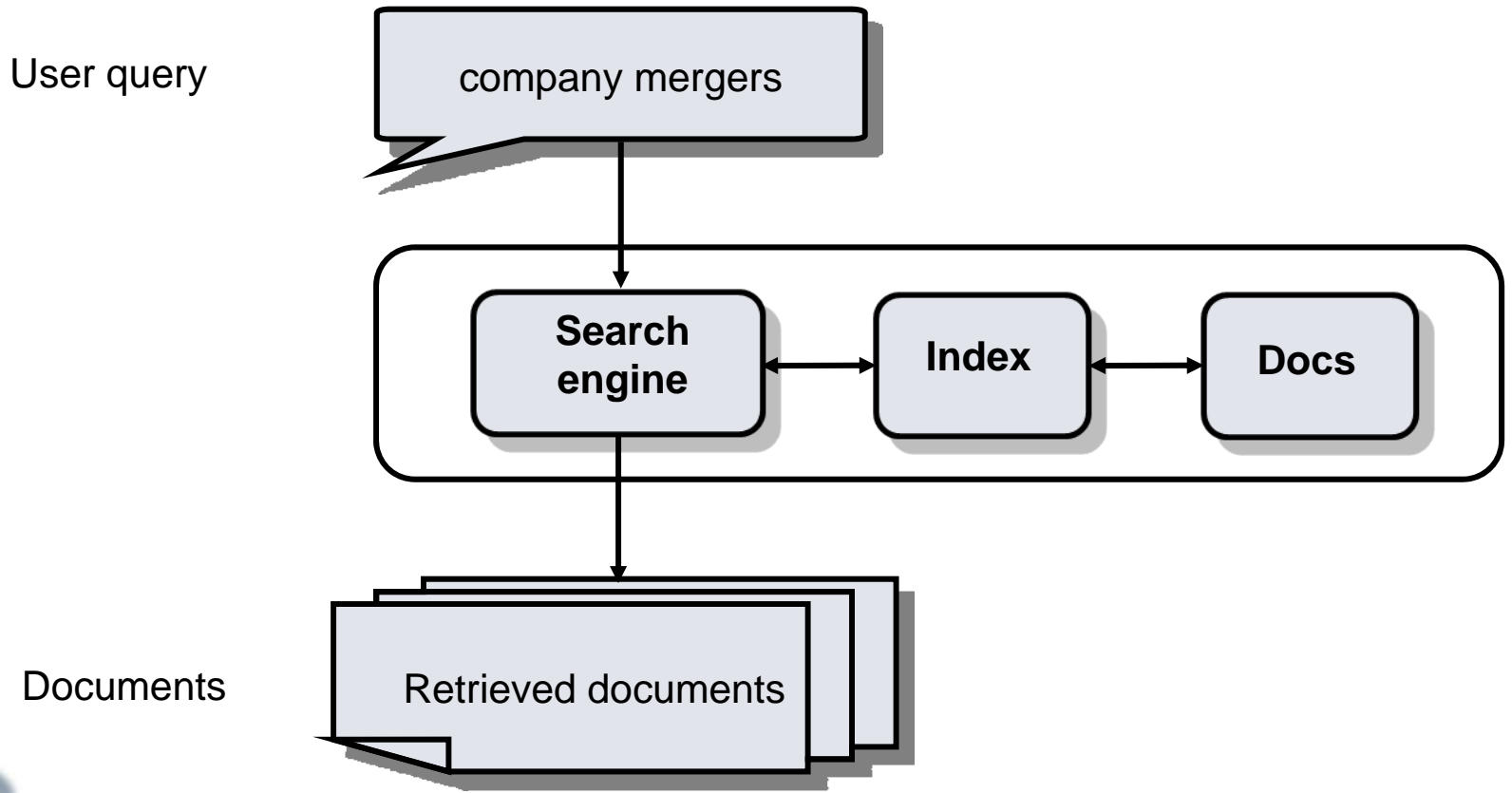
## NaturalFinder - Integration

- **It can easily be integrated with any search engine**
  - dtSearch
  - Google Search Appliance
  - Lucene
  - Memex
  - Live Search...
- Guaranteed by
  - Architecture - top layer
  - Technical approach - libraries & text-based API



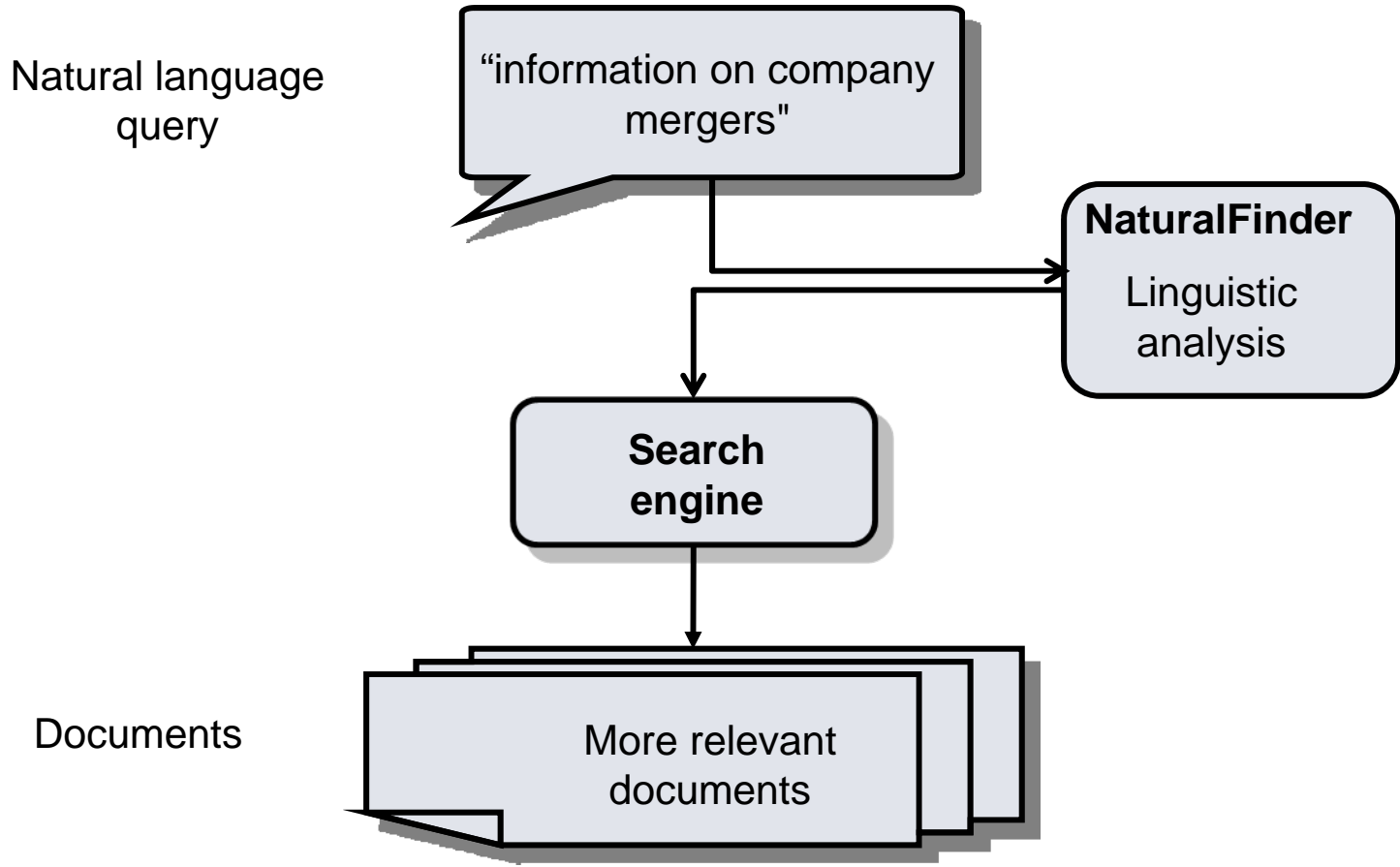
# NaturalFinder - Integration

Architecture of a search engine



# NaturalFinder - Integration

Top layer between user interface and search engine



## NaturalFinder - Integration

- Technical features
  - Proprietary development: 100% in C/C++
  - Multiplatform: Windows, Linux/UNIX
  - Modular components: spelling, stemming...
  - Flexible text-based API
- Performance (in Intel Pentium 4, 2.8GHz)
  - Lexical processing: 1,000,000 words per second
  - Syntactic processing: 1,000 sentences per second
- With any application and any platform

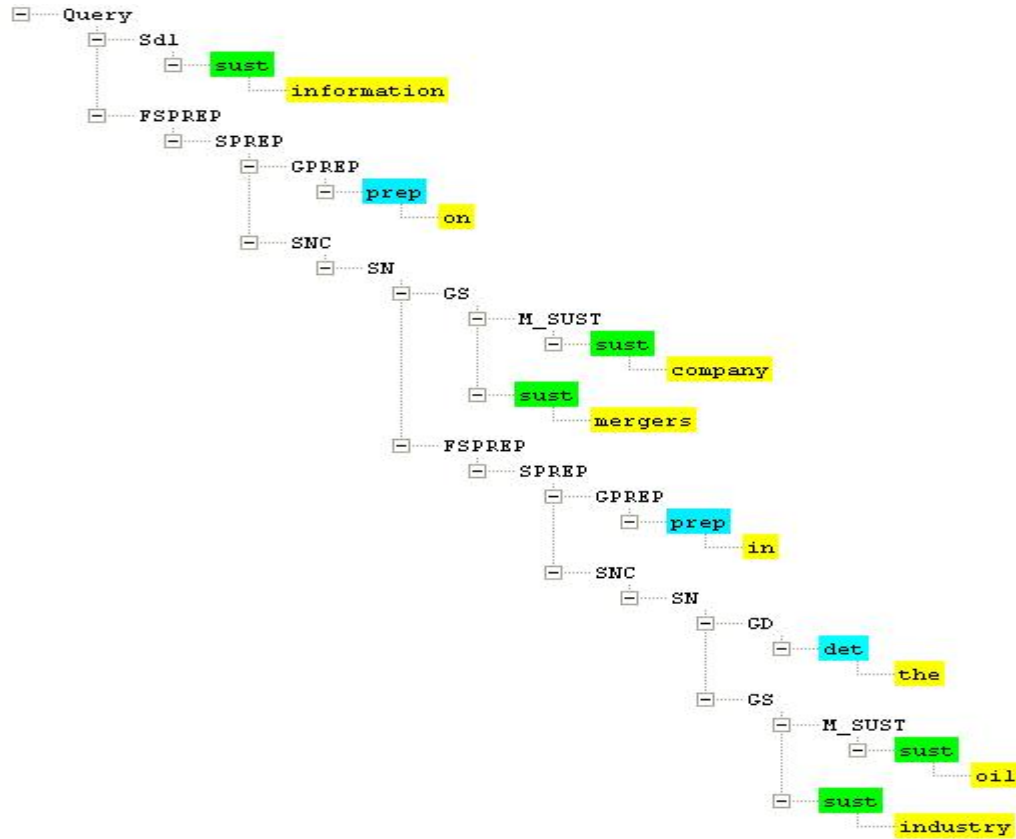


## NaturalFinder - Technology

- DataSuite - Suite of components
- DataGrammar: natural language parser
  - *information on Paris but not Hilton*
- DataLexica: lexical database (POS, morphology)
  - *corpus = corpora*
  - *new < > news*
- DataSpell: spell checker
  - *recentnews > recent news*
  - *data base > database*
- DataNet: semantic relationship management
  - *<Rome> <PART OF> <Italy>*
  - *<Rome> <TYPE OF> <CITY>*
  - *<Rome> <POPULATION> <2,656,000>*



# NaturalFinder - DataSuite syntactic tree



*Information on company mergers in the oil industry*



## Our proposal - In short

- Bitext, provider of natural language technologies
  - Robust and flexible
  - Multipurpose
- Action plan proposed
  - Identify needs
  - Adapt solution
  - Build agreement
    - NDA
    - Technical issues
    - Business issues
    - Cooperation agreement



# Thank you

Antonio S. Valderrábanos  
Founder & CEO  
Bitext - The Bits and Text Company  
Edificio Prisma, 1, 1  
Calle Cólquide 6  
28230 Las Rozas Madrid  
Tel: +34 911461660

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

