

# Bitext - The Bits and Text Company

Natural language technologies  
and user-generated content:  
NaturalOpinions

Bitext - The Bits and Text Company  
info@bitext.com  
<http://www.bitext.com>  
91.146.16.60



*Bitext.com*  
THE BITS AND TEXT COMPANY

## Index

- Information extraction in user-generated content
- Opinion analysis for business intelligence
- Natural language technologies and information extraction
  
- NaturalOpinions for English
- NaturalOpinions for Spanish
  
- Bitext: search engine and content analysis industry



## User-generated content

- User-generated content:  
Chats, forums, RSS, applications, podcasts, social media and blogs
- Continuous growth of number of blogs
  - Technorati keeps track of 112.8 million blogs (February 2008)
- Ever-growing number of consumers make comments on brands and products in blogs and posts
- Studies show that consumers change their mind after reading blogs with comments about a product
- Business Intelligence: institutions need to know what people are saying about a brand/product in order to define marketing strategies, improve products, etc.



## Opinion analysis for business intelligence

- Bitext's approach: natural language technologies
  - NaturalOpinions, especially customized to understand the language used in blogs, Twitter, etc.
- Opinion polarity determination
  - Entity: list of brand names in dictionaries
  - Entity attributes: determined by grammatical rules
  - Voice: what is said about the entities or attributes
  - Polarity: direct or negated
- Result: opinions classified according to
  - Relevance, Polarity, Voice, and Attributes



## Natural language technologies and information extraction

- DataSuite:
  - SDK for natural language processing
  - Rule-based processing, not based on statistics
  - Available in Spanish, English, French, Catalan, and Basque
- DataSuite consists of:
  - DataLexica: morphological information
  - DataGrammar: syntactic/semantic analyzer
  - DataSpell: spell-checker
  - DataNet: semantic relations



## Natural language technologies and information extraction

- Technical features of DataSuite:
  - Developed in C/C++ for Linux/UNIX and Windows
  - Flexible API
  - Integration with third-party applications
- DataSuite's performance (Intel Pentium 4, 2.8GHz):
  - Lexical processing: 1,000,000 words per second
  - Syntactic processing: 1,000 sentences per second



## Examples of opinion polarity determination

- **Example in Spanish:** "...por eso nunca me han gustado los anuncios de Nike"
- **Output of DataSuite:**
  - entity: "Nike"
  - entity-type: "BRAND"
  - entity-component: "anuncios"
  - component-attribute: "IMAGE"
  - polarity: "NEGATED"
  - voice: "nunca me han gustado"
  - verb: "me han gustado"
  - direct-complement: "los anuncios de Nike"
  - adverbial-complement: "por eso"
  - adverbial-complement: "nunca"



# NaturalOpinions for Spanish


NaturalOpinions

por eso nunca me han gustado los anuncios de Nike

Analyze Sentence    Next Sentence    Load Corpus    EXIT

```
<opinion-block>
<sentence>"por eso nunca me han gustado los anuncios
de Nike"</sentence>
<entity>"Nike"</entity>
<entity-referenced>"NO"</entity-referenced>
<entity-type>"BRAND"</entity-type>
<entity-component>"los anuncios"</entity-component>
<entity-component-norm>"los[det,el] anuncios
[sust,anuncio]"</entity-component-norm>
<entity-component-referenced>"NO"</entity-component-
referenced>
<component-attribute>"IMAGE"</component-attribute>
<polarity>"DIRECT"</polarity>
<opinion>"nunca me han gustado"</opinion>
<opinion-norm>"nunca[adv,nunca] me[adv,me] han
[verb,haber] gustado[verb,gustar]"</opinion-norm>
</opinion-block>
```

NVOID  
GS\_NED  
POLARITY-DIRECT-IMAGE-BRAND  
VOICE-direct  
SADV  
adv  
nunca  
Voicetype3tr  
adv  
me  
Verbs  
Verbsref  
verb  
han  
verb  
gustado  
ATTR-brand-IMAGE  
Brand-image-PART-OF  
GDet  
det  
los  
sust  
anuncios  
GPrep  
unknown  
de  
SN-ENTITY-brand  
sust  
Nike

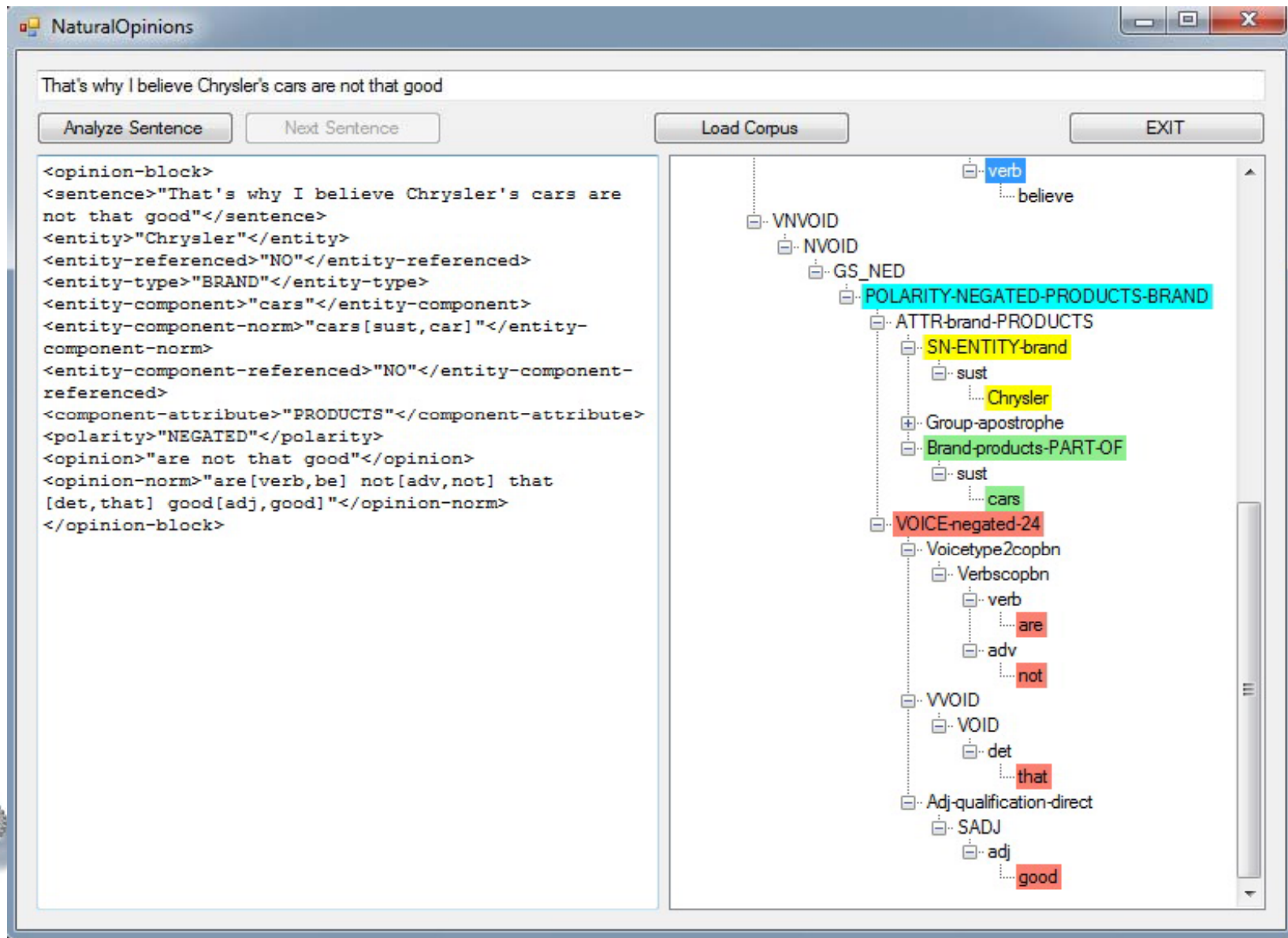


## Examples of opinion polarity determination

- **Example in English:** "...That's is why I believe Chrysler's cars are not that good"
- **Output of DataSuite:**
  - entity: "Chrysler"
  - entity-type: "BRAND"
  - entity-component: "cars"
  - component-attribute: "PRODUCT"
  - polarity: "DIRECT"
  - voice: "are not that good"
  - verb: "are not"
  - predicate-complement: "that good"



# NaturalOpinions for English



The image shows the NaturalOpinions software interface. The window title is "NaturalOpinions". The main text area contains the sentence: "That's why I believe Chrysler's cars are not that good". Below the text are four buttons: "Analyze Sentence", "Next Sentence", "Load Corpus", and "EXIT".

The left pane displays the XML output of the analysis:

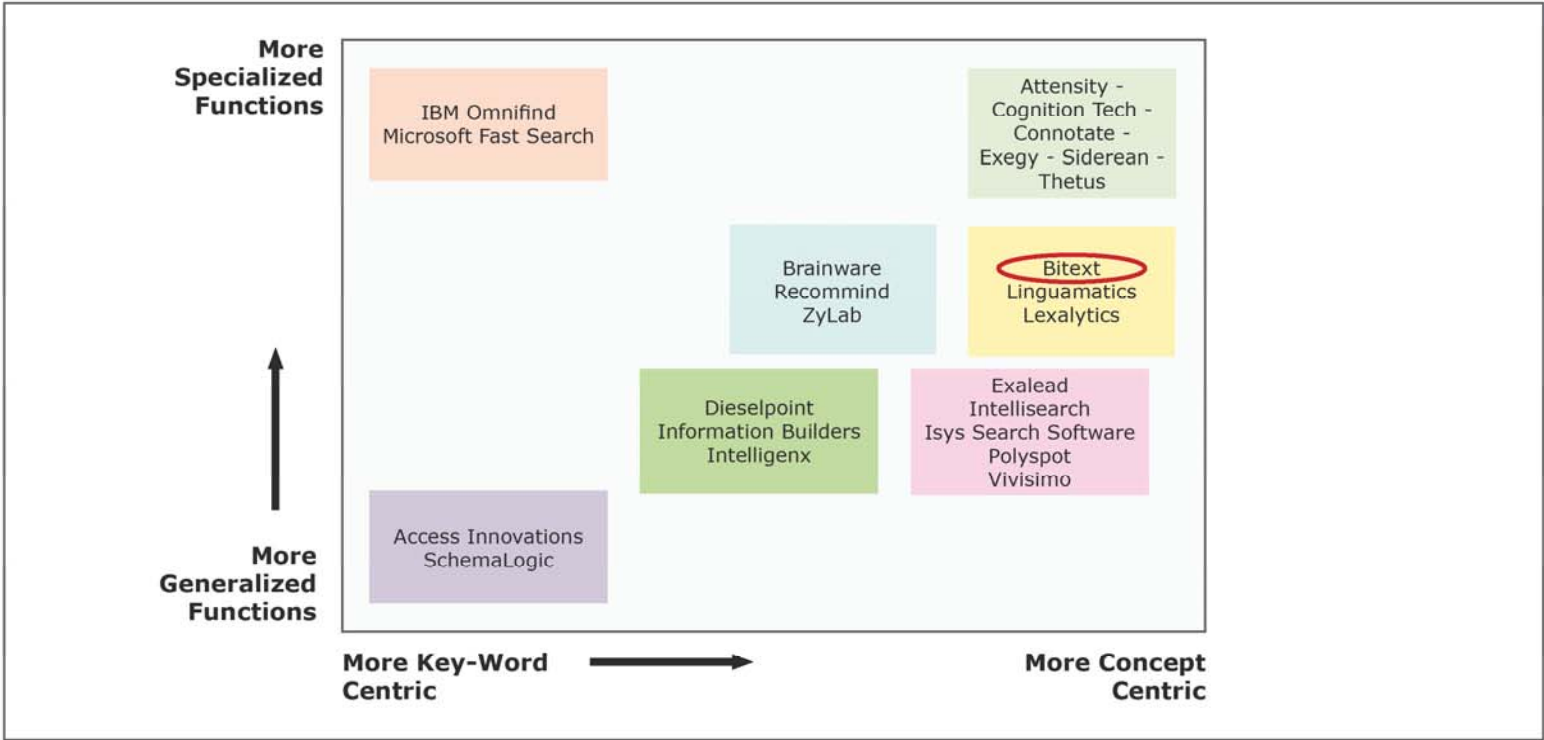
```
<opinion-block>
<sentence>"That's why I believe Chrysler's cars are
not that good"</sentence>
<entity>"Chrysler"</entity>
<entity-referenced>"NO"</entity-referenced>
<entity-type>"BRAND"</entity-type>
<entity-component>"cars"</entity-component>
<entity-component-norm>"cars [sust, car]"</entity-
component-norm>
<entity-component-referenced>"NO"</entity-component-
referenced>
<component-attribute>"PRODUCTS"</component-attribute>
<polarity>"NEGATED"</polarity>
<opinion>"are not that good"</opinion>
<opinion-norm>"are [verb, be] not [adv, not] that
[det, that] good [adj, good]"</opinion-norm>
</opinion-block>
```

The right pane displays a hierarchical tree diagram of the sentence structure. The root node is "verb" (believe). The tree structure is as follows:

- verb (believe)
- VNVOID
- NVOID
- GS\_NED
- POLARITY-NEGATED-PRODUCTS-BRAND
- ATTR-brand-PRODUCTS
- SN-ENTITY-brand
- sust (Chrysler)
- Group-apostrophe
- Brand-products-PART-OF
- sust (cars)
- VOICE-negated-24
- Voicetype2copbn
- Verbscopbn
- verb (are)
- adv (not)
- VOID
- VOID
- det (that)
- Adj-qualification-direct
- SADJ
- adj (good)



# Bitext: search engine and content analysis industry



Quadrant for Market Sector Vendors

Gilbane Report – “Beyond Search – What to do when your search engine does not work”



# Thank you!

Bitext - The Bits and Text Company  
Edificio Prisma, 1, 1  
Calle Cólquide 6  
28230 Las Rozas Madrid - SPAIN  
Phone: +34 911461660

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

