

Natural Language and Text Processing Laboratory

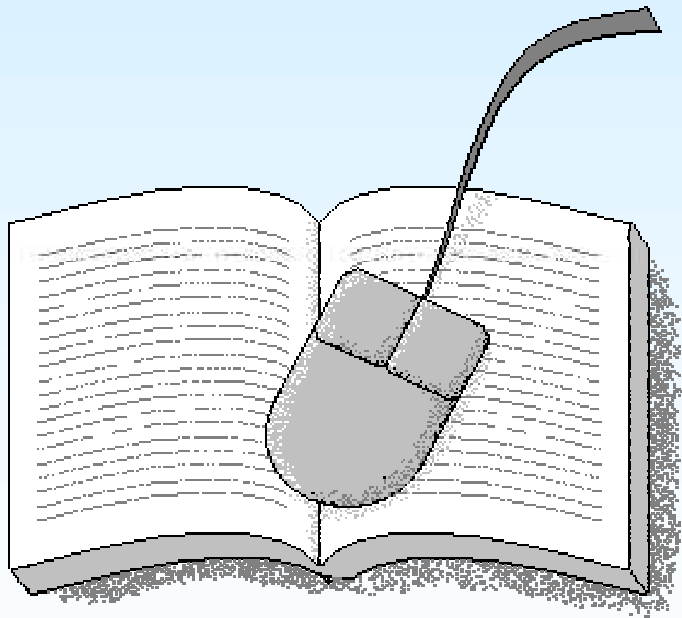
Projects and Research Directions

Head: *Alexander Gelbukh*
gelbukh@cic.ipn.mx

Contents

- General info
- Main research directions
- Projects in Computational linguistics
- Projects in Text processing applications
- Projects in Mathematical modeling
- International contacts
- Possible future projects

NLP&TP Lab



- Since 1996
- 4 doctors
- 4 doctoral students
- 3 master students
- 23 projects
- 100+ publications
- 1 International Conference

Projects

- 5 CONACyT (1 Joint w/ Moscow City Government)
- 2 REDII-CONACyT
- 13 CGEPI-IPN (1 Joint w/ Mexican Oil Institute – IMP)
- 1 Senate of Mexican Republic
- 1 Joint w/ Moscow City Government

Main Research Directions

- Theoretical Computational Linguistics
 - Syntax analysis
 - Semantic analysis
 - Dictionaries and resources
- Intelligent Text Processing Applications
 - Information retrieval
 - Classification and text mining
 - Topical summarization
- Applied Mathematics
 - Modeling of diffusion processes

Projects in Computational Linguistics

- **CONACyT 26424-A:**
Advanced syntactic analyzer for Spanish
- **REDII-CONACyT:**
Subcategorization dictionary for Spanish
- **CONACyT 32003-A:**
Compilation of text corpus of a new type through Internet
- **CONACyT J31989:**
Extraction of semantic info from bilingual dictionaries

Parser [input.txt : analyze words]

File Help

Parser icons: Folder, Document, Checkmark, Arrow down, Bar chart, Hash, C, D, G, G, ?

Without our tool

Text: Veo gato con telescopio.

Trees Morphology Dump Tracing PMS

Parse tree structure:

- S
 - VP_no_needs_NP(SG,1PRS,MEAN)
 - V(SG,1PRS,MEAN) → Veo
 - NOM(SG,MASC)
 - N(SG,MASC) → gato
 - PP
 - PR → con
 - N(SG,MASC) → telescopio
 - \$PERIOD → .

1	1.9e-15
2	1.9e-15
3	4.6e-12
4	4.6e-12
5	4.6e-12
6	4.6e-12

Previous Next

Font height: 17

Total variants 6

Parser [input.txt : analyze words]

File Help

C D G

Text

Veo gato con telescopio.

Trees Morphology Dump Tracing PMS

Syntax tree diagram for the sentence "Veo gato con telescopio.":

```

    S
    / | \
    VP VP NP
    / | \ | \
    VP N PR N
    / | \ | \
    V N PR N
    | | | |
    Veo gato con telescopio
  
```

VP_no_needs_NP(SG,1PRS,MI\$PERIOD)

VP_no_needs_NP(SG,1PR\$PP\$MEAN)

V(SG,1PRS,MEAN) N(SG,MASC) PR N(SG,MASC)

Veo gato con telescopio .

3	4.6e-12
4	4.6e-12
5	4.6e-12
6	4.6e-12
1	1.9e-15
2	1.9e-15

Previous Next

Previous Next

Font height: 17

Total variants 6

With our tool

Projects in Applied Text Processing 1

- REDII-CONACyT (joint with ITESM):
Agent-based search engine for digital libraries
- Senate of Mexican Republic:
Concept-based search engine for the DB of the Senate

Projects in Applied Text Processing 2

- CONACyT & Moscow City Government,
No. 0402-00:
Investigation of large ecological document bases
- Joint with Moscow City Government:
Knowledge discovery in the flows of the letters
of city dwellers

Topical summary

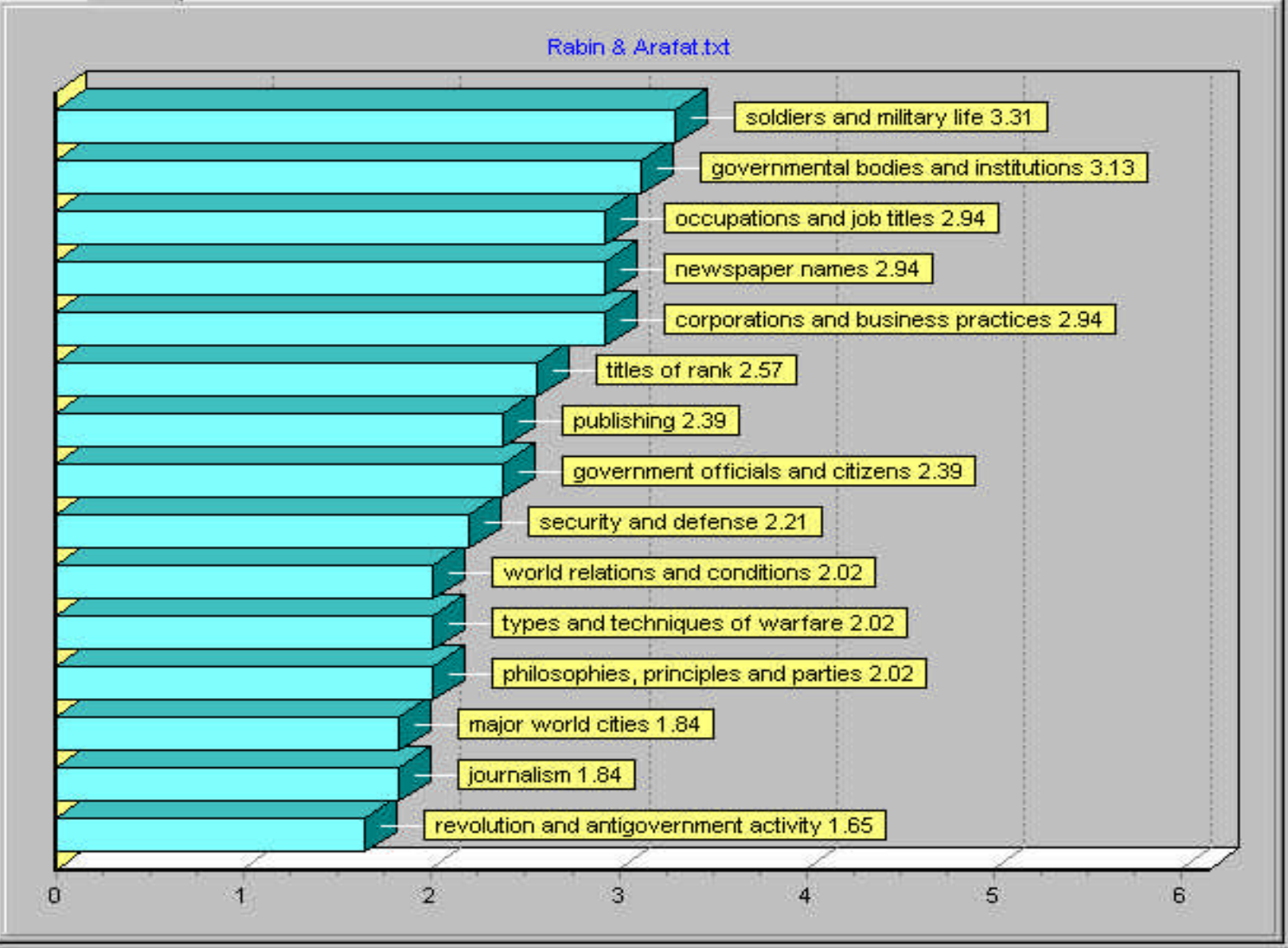


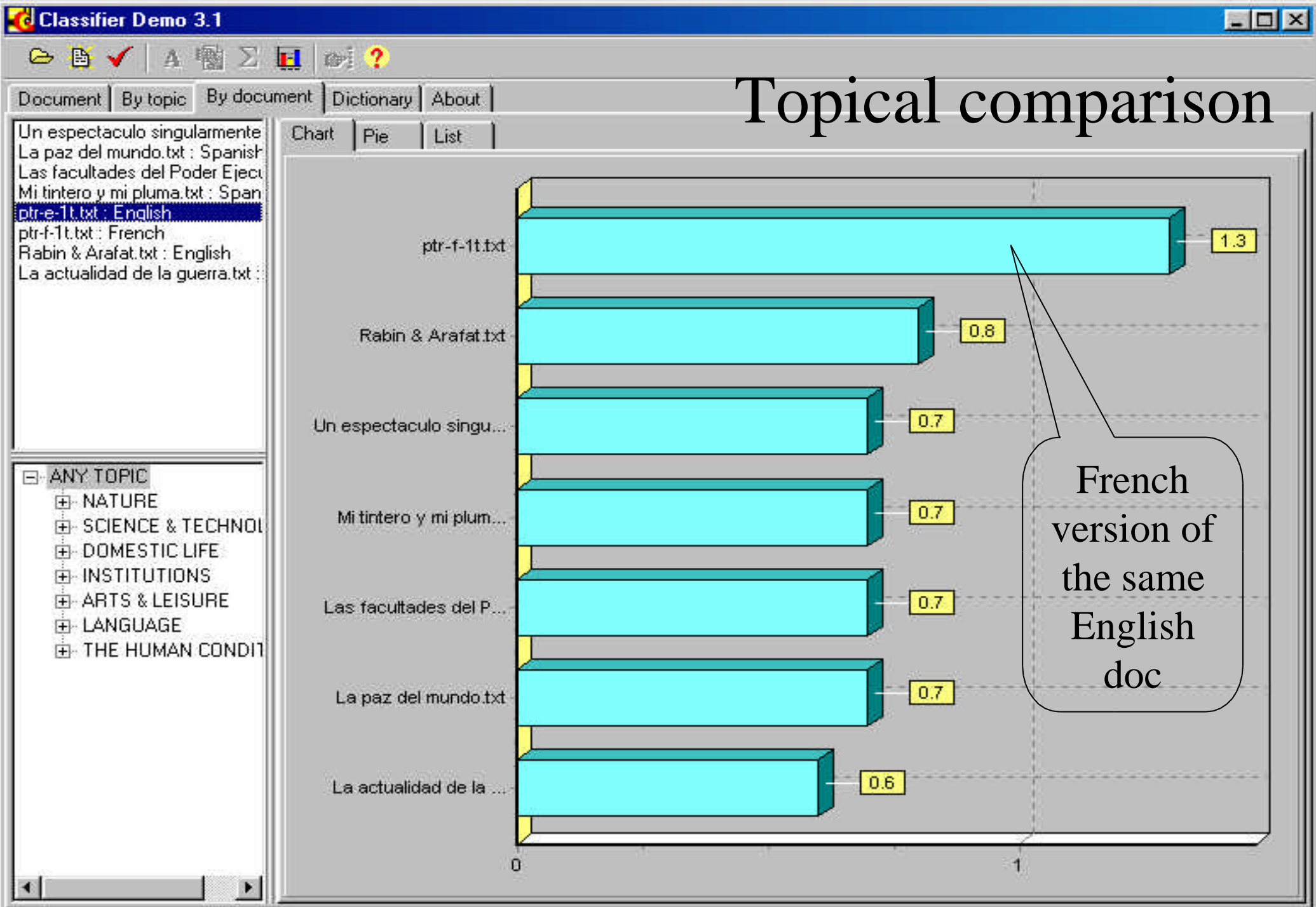
Document | By topic | By document | Dictionary | About

Un espectáculo singularmente
 La paz del mundo.txt : Spanish
 Las facultades del Poder Eject
 Mi tintero y mi pluma.txt : Span
 ptr-e-1t.txt : English
 ptr-f-1t.txt : French
Rabin & Arafat.txt : English
 La actualidad de la guerra.txt :

Text | Chart | Pie | List

- [-] ANY TOPIC
 - [+] NATURE
 - [+] SCIENCE & TECHNOLOGY
 - [+] DOMESTIC LIFE
 - [+] INSTITUTIONS
 - [+] ARTS & LEISURE
 - [+] LANGUAGE
 - [+] THE HUMAN CONDITION





Dictionary used

- storms
- overcast weather
- clear weather
- clouds
- hot weather
- cold weather
- mild weather
- wet weather
- forecasting and meteorology
- [-] SCIENCE & TECHNOLOGY
 - science and technology
 - [-] THE SCIENCES
 - disciplines and specialties
 - [-] PHYSICS
 - physics
 - nuclear and particle physics
 - principles of mechanics, waves and measurement
 - branches and disciplines of physics
 - substances, particles and atomic architecture
 - optics
 - cosmology
 - [-] ASTRONOMY & SPACE SCIENCE
 - the science
 - astronomy and space science
 - celestial phenomena and points
 - celestial bodies
 - space exploration and rocketry
 - [-] MATHEMATICS
 - mathematics
 - mathematical tools
 - geometric shapes and mathematically defined forms

- élément du transuranic
- éleveur réacteur
- émission secondaire
- énergie de la séparation
- énergie obligatoire
- énergie obligatoire nucléaire
- équation de la force forte
- équation nucléaire
- accélérateur
- accélérateur linéaire
- albedo
- amu
- antimatière
- antiparticule
- antiquark
- atome
- atome du rutherford
- atome-qui fend
- backscatter
- baryon
- boson
- briseur de l atome
- calutron
- carbone
- centrifugeuse du gaz
- chambre de la bulle
- chambre du nuage
- charme
- coeur
- collider

- English
- Spanish
- French

Projects in Mathematical Modeling

- CONACyT 27928-A:
Modeling of air pollution in Mexico City
- Joint with IMP, No. 990080 by CGEPI-IPN:
Detection of local pollution sources



Titles

The easy and fast way to regulatory air modeling

[Optional Second Title]



Dispersion Options

- Regulatory Default
- Non-Default Options

Output Types

- Concentration
- Total Deposition (Dry and Wet)
- Dry Deposition
- Wet Deposition

Run Option

- Run
- Do Not Run

Control Pathway [X]

In the Control Pathway you define: dispersion options, dispersion coefficient, pollutant, averaging period, etc.

We are adding functionality to this feature

Non-Default Options

- Gradual plume rise
- No buoyancy-induced dispersion
- No stack-tip downwash
- Missing data processing routine
- Bypass the calms processing routine

Dispersion Coefficient

- Rural
- Urban

Plume Depletion Due To

- Dry Removal
- Wet Removal

Pollutant/Avg Time/...

Optional Files

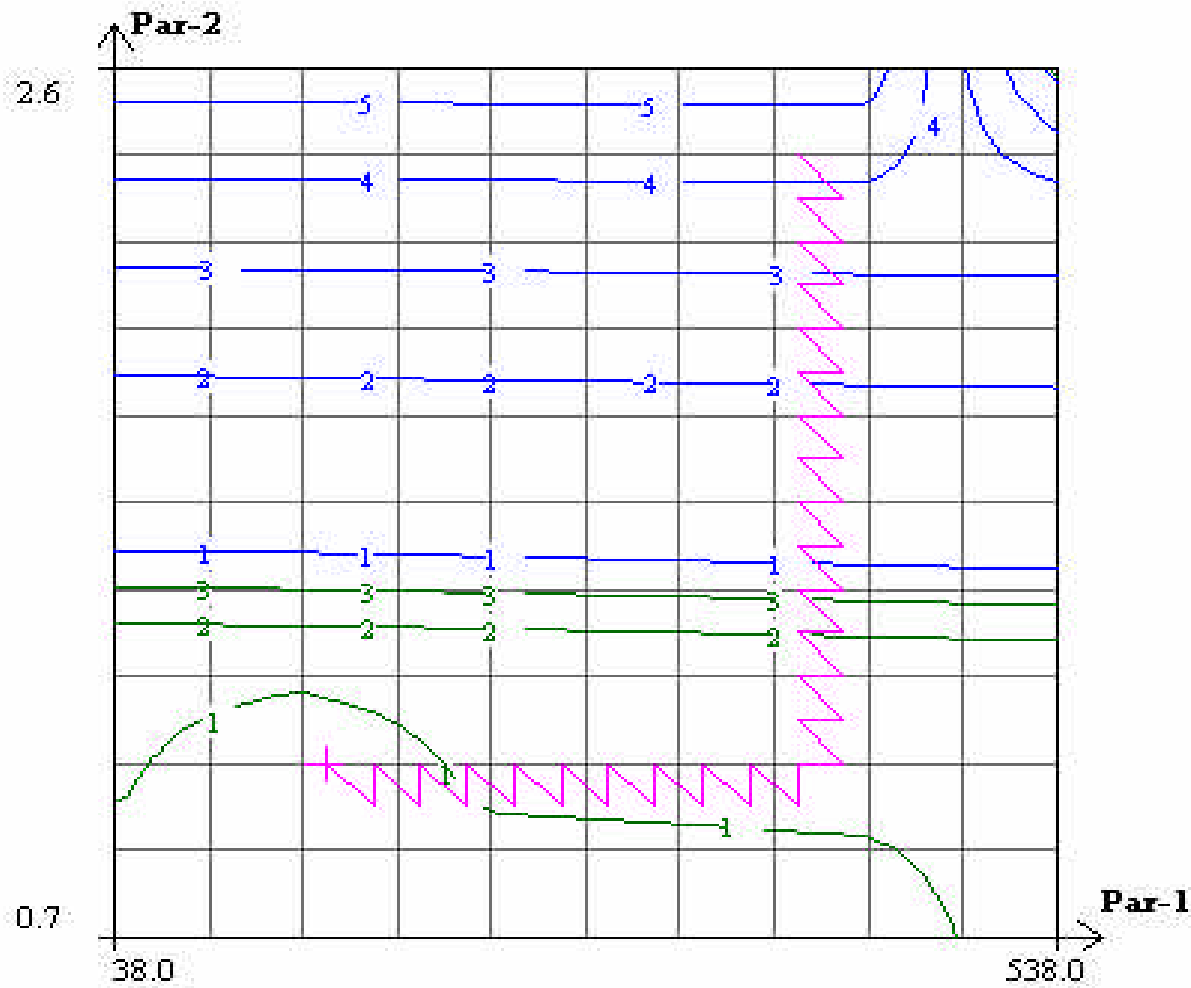
Cancel

Next >

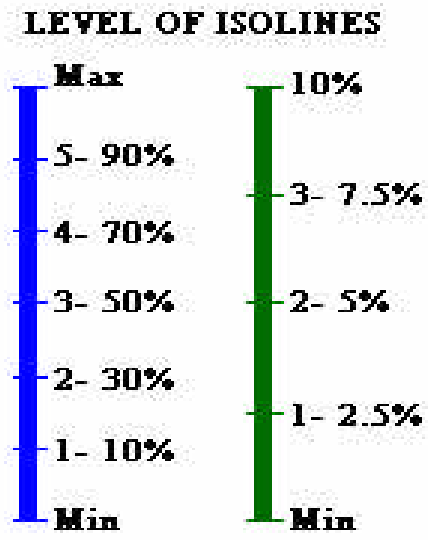




TRAJECTORY OF OPTIMIZATION PROCESS



Method of Nelder-Mead
2D Problem



International Contacts

- CYTED and RITOS2
(Euro-American Research Organizations):
Member group
- City Government, Moscow:
Common Projects
- UPC, Barcelona:
Student exchange & research contacts
- U. Montreal:
Research in Meaning-Text Theory

Possible Future Joint Projects

- Heymans Institute, Gröningen, Holland:
Classification of complex interdisciplinary information in very large databases
- City Government, Moscow:
Decision making on the basis of laws and normative deeds
- U.P.C., Barcelona:
Spanish in HPSG



Thank you!