



Analiza automată a discursului științific

-Teorie, programe, aplicații -

Prof. Dr. Dan CARAGEA



Proiect finanțat de
UNIUNEA EUROPEANĂ



MINISTERUL MUNCII
FAMILIEI ȘI PROTECȚIEI SOCIALE
AMPOSDRU



FONDUL SOCIAL EUROPEAN
POSDRU
2007-2013



INSTRUMENTE STRUCTURALE
2007-2013



MINISTERUL EDUCAȚIEI
CERCETĂRII ȘI INOVĂRII
OIPOSDRU



UNITATEA EXECUTIVĂ PENTRU
FINANȚAREA ÎNVĂȚĂMÂNTULUI
SUPERIOR ȘI A CERCETĂRII
ȘTIINȚIFICE UNIVERSITARE



Preliminarii

În activitatea de cercetare și de elaborare a discursurilor științifice, un număr tot mai mare de specialiști (dar și studenți) folosesc programe speciale (*software*-uri) pentru **automatizarea** unor etape ale muncii lor:

- **colectarea** de informații (articole, studii) de pe Web cu ajutorul unor roboți (*spider*);
- crearea unor **scenarii** (arborele conceptual)
- **filtrarea** fișierelor colectate pentru a ajunge la informația cu adevărat pertinentă pentru cercetător;
- **analiza** unor texte care pot suscita un interes special (*text mining*);
- **autoanaliza** discursului produs: compararea text/rezumat, extragerea cuvintelor cheie;
- posibilitatea **publicării on-line** a unor rezultate, clasificări etc.



Analiza discursului

Analiza discursului este un demers multidiscplinar, calitativ și cantitativ, care studiază **conținutul** și **contextul** discursului (oral și scris).

Analiza discursului se sprijină pe **concepte împrumutate** din retorică, lingvistică, științele comunicării, psihologie, sociologie, istorie, informatică etc.

Pentru unii specialiști, discursul este considerat o **realitate în sine**, și nu un reflex al realității, așa cum se afirmă în analiza conținutului.

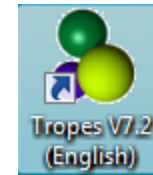
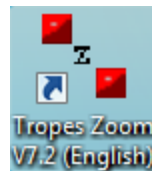
Analiza evidențiază o serie de realități textuale care sunt **relații de putere, de incluziune, excluziune...** dar și o **coerență** (referențială, structurală și argumentativă).

Unele etape ale analizei discursului pot fi automatizate, dar analiza continuă să fie un produs eminentemente uman (**analiză și interpretare**)



Gama Tropes-Zoom Web

TROPES ZOOM cu **MODUL WEB** (fabricant: **Acetic**, Franța – www.acetic.fr sau www.semantic-knowledge.com) este gama de programe propusă spre ilustrare fiind, foarte probabil, una dintre cele mai interesante de pe piața mondială.



- **Robot V1.0** – *Spider* care permite colectarea informației de pe Web.
- **Zoom V7.2** – Motor de căutare, de clasificare și filtrare semantică, dotat cu un nucleu de indexare în limbaj natural.
- **Tropes V7.2** – *Software* de analiză a textelor: clasificare semantică, analiză calitativă, analiza discursului, extragerea de cuvinte cheie; construcția de *thesauri* și ontologii, publicarea *on-line* a rezultatelor.



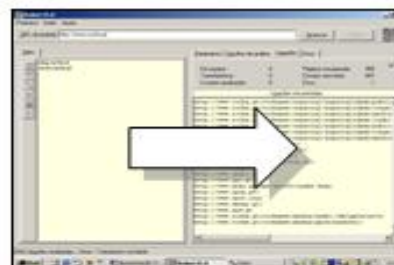
Gama Tropes-Zoom Web



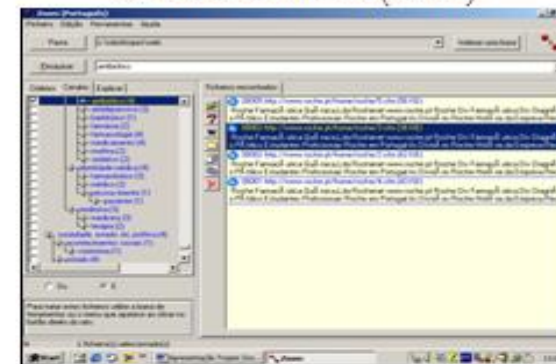
Identificarea surselor informative



Colectarea



Cautarea și filtrarea (Zoom)

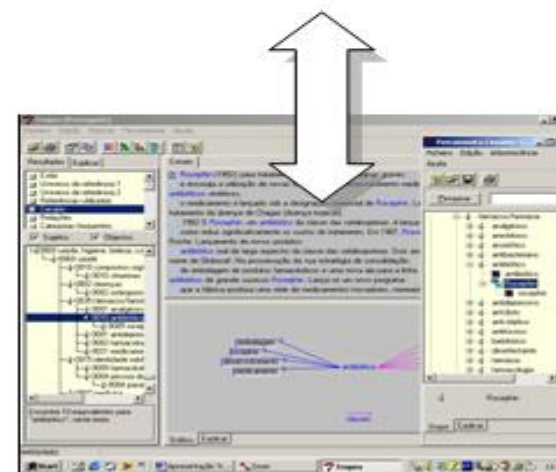


(Robot)
Tropes Zoom: secvența
completă de resurse
de analiză a conținuturilor

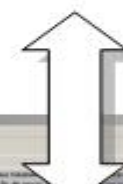
Generarea de conținuturi



(Modul Web)



Analiza și clasificarea (Tropes)



Publicarea (Internet sau Intranet)



Fundamentul teoretic

Tropes de bazează pe **Analiza cognitivo-discursivă** (Ghiglione, Rodolphe și Blanchet, A. – *L'Analyse cognitivo-discursive*, PUG, Grenoble, 1995) și care presupune:

- decuparea textului în propoziții (unități cognitivo-discursive);
- clasificarea semantică a cuvintelor utile;
- identificarea claselor paradigmatică de substantive (**referințe nodale**);
- obținerea unui model propozițional simplificat din punct de vedere formal (**structura fundamentală de semnificare**).

Rezumând, este vorba de extragerea din text a unei serii de variabile, tratate apoi statistic, pentru a obține rezultate obiective de analiză (identificabile sau nu la o lectură mai profundă a textului).



Funcționarea internă

- Lanțuri de caractere >> ANALIZA MORFOLOGICĂ
- Forme lexicale >> ANALIZA LEXICALĂ
- Frazе >> ANALIZA SINTACTICĂ
- Cuvinte, propoziții >> ANALIZA SEMANTICĂ
- Sensuri >> ANALIZA COGNITIVO-DISCURSIVĂ
- SFS >> STATISTICI



Categoriile gramaticale

- **SUBSTANTIVE** (comune și proprii)
- **ADJECTIVE** (subiective, obiective, numerice)
- **VERBE** (factive, stative, declarative, performative)
- **PRONUME** (personale; altele)
- **MODALIZARI** (timp, loc, mod, afirmație, îndoială, negație, intensitate)
- **CONNECTORI** (condiție, cauză, scop, adiție, disjuncție, opoziție, comparație, timp, loc)



Rețele Semantice

- REFERINȚE UTILIZATE («București, Cluj, Iași»)
- UNIVERSUL DE REFERINȚĂ 2 («România»)
- UNIVERSUL DE REFERINȚĂ 1 («Geografie»)

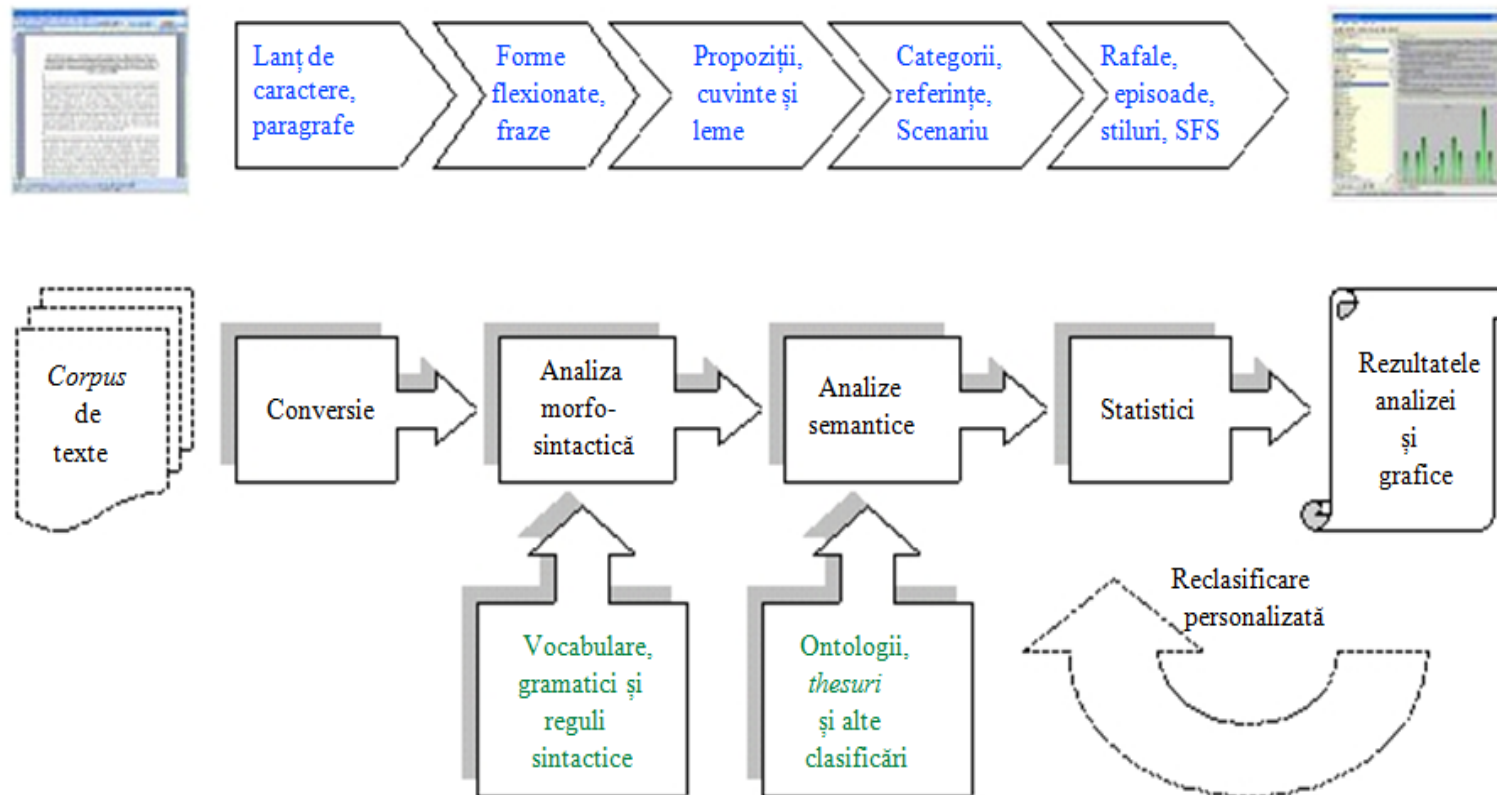


Structura Fundamentală de Semnificare

- **Relații** (inter-referențiale)
- **Episoade și rafale**
- **Fraze remarcabile** (triaj)
- **Stil**
- **Statistici**



Procesul de analiza





Caracteristici

- Metode confirmate, bazate pe **cercetarea universitară**.
- O preocupare permanentă pentru **obiectivitate**: **Tropes** își «justifică» analizele, prezentând indicatorii luați în considerare în diagnosticul său.
- O **ergonomie** simplă, pentru un **software interactiv**, destinat analiștilor de orice nivel, mai ales celor obișnuiți cu suportul hârtie.
- O logica de **inteligență artificială** capabilă să rezolve ambiguitățile (lexicale și semantice) și să clasifice majoritatea cuvintelor unei limbi.
- **Dicționare enciclopedice pragmatice**: suportă peste 500.000 de clasificări semantice, inclusiv mărci și nume de firme...
- Un motor de analiză lingvistică **foarte rapid**, capabil să trateze (în fiecare zi) mii de documente într-un computer obișnuit.



Aplicații

Gestiunea Informației și a Cunoașterii

Indexarea și clasificarea bazelor documentare. Crearea de *thesauri*.

Marketing, sondaje și chestionare

Analiza calitativă/cantitativă a datelor textuale. Studiul concurenței. Analiza presei.

Comunicarea în Relații Publice, Marketing

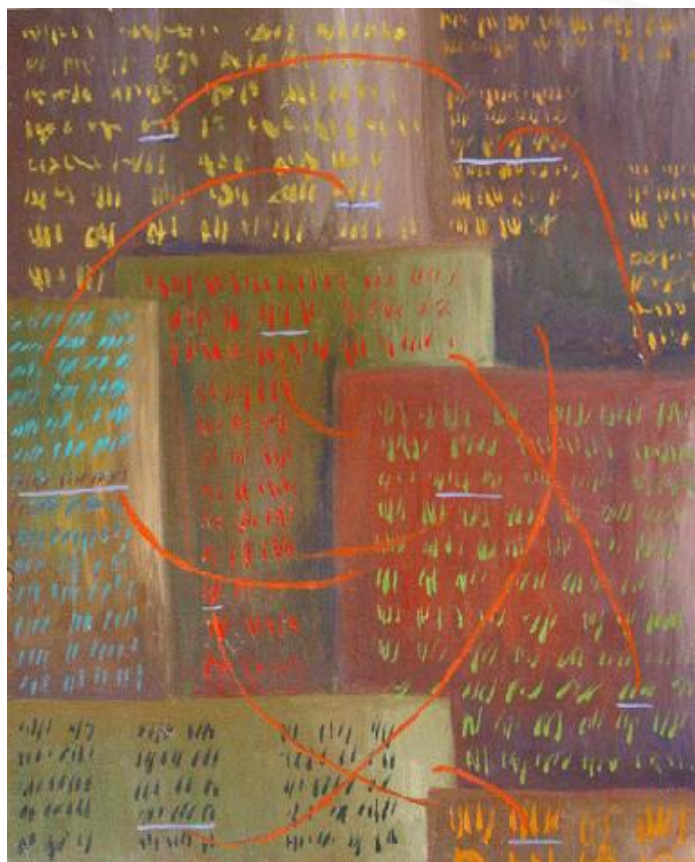
Studiul imaginii de marca al unei firme. Analiza mesajelor strategice. Analiza *mailing*-ului.

Alte domenii

Inteligență economică. Informatică decizională. Studii sociologice. Analiza discursului politic. Învățământ.



Coordonatele noastre



Bel Llodrà Riera, Hyperlinks

Dan Caragea

UEFISCSU

dan.caragea@uefiscsu.ro