UNIUNEA EUROPEANĂ
Fondul Social European

GUVERNUL ROMÂNIEI
Ministerul Dezvoltării Regionale
și Administrației Publice

PODCA
INOVAȚIE ÎN ADMINISTRAȚIE

Instrumente Structurale
2007-2013

Proiect cofinanţat din Fondul Social European, prin Programul Operaţional "Dezvoltarea Capacităţii Administrative", în perioada 2007-2013

Dezvoltarea platformei de management de cunoaștere

UEFISCDI
UNITATEA EXECUTIVĂ PENTRU FINANȚAREA ÎNVĂȚĂMÂNTULUI SUPERIOR, A CERCETĂRII, DEZVOLTĂRII ȘI INOVĂRII

Creşterea Capacităţii Administrative a Sistemului Public de CDI

UNIUNEA EUROPEANĂ
Fondul Social European

GUVERNUL ROMÂNIEI
Ministerul Dezvoltării Regionale
și Administrației Publice

PODCA
INOVAȚIE ÎN ADMINISTRAȚIE

Instrumente Structurale
2007-2013

Proiect cofinanţat din Fondul Social European, prin Programul Operaţional "Dezvoltarea Capacităţii Administrative", în perioada 2007-2013

Dezvoltarea platformei de management de cunoaștere

Cuprins

UEFISCDI
UNITATEA EXECUTIVĂ PENTRU FINANȚAREA ÎNVĂȚĂMÂNTULUI SUPERIOR, A CERCETĂRII, DEZVOLTĂRII ȘI INOVĂRII

Creşterea Capacităţii Administrative a Sistemului Public de CDI

Platforma de management de cunoastere isi propune sa deserveasca structurile MENCS responsabile în domeniul CDI precum şi UEFISCDI, principala institutie cu rol in finantarea cercetarii din Romania. Mecanismele dezvoltate in proiect pot sustine orientarea strategica a organizatiilor cu activitate de cercetare. De asemenea, module ale acestei platforme pot contribui la consultari multi-actor pe teme de CDI, precum si la elaborarea politicilor CDI bazate pe evidente. Premisa este că de înţelegerea tehnologiilor emergente – tehnologii cu potenţial de adoptare pe scară largă şi/sau impact major asupra unuia sau mai multor sectoare economice – depinde capacitatea actorilor din ecosistemul de cercetare, dezvoltare, inovare de a îşi construi strategii şi planuri pe termen lung. Identificarea din timp a tehnologiilor relevante va creşte competitivitatea actorilor amintiţi şi capacitatea lor strategică.

Platforma de management de cunoastere se constituie ca un „radar" de identificare a tendintelor tehnologice - un sistem inovativ care combina analisti umani si inteligenta artificiala pentru selectarea dintr-un set larg de stiri a acelora care reprezinta "semnale slabe", respectiv care descriu fenomene emergente (cu posibilitate de amplificare), in special din domeniile tehnologice. Acest demers a presupus dezvoltarea unei baze cu stiri, preluate din peste 300 de surse online, care in acest moment inglobeaza peste 650.000 de stiri unice din perioada 2010-2015. Procesul de identificare a fenomenelor emergente are doua componente: o componenta de validare umana si o componenta automata care foloseste algoritmi de inteligenta artificiala si tehnici de procesare a limbajului natural. Rezultatele celor doua componente majore, precum si restul stirilor din repozitoriu au fost integrate intr-un modul software de vizualizare multicriteriala (**R7.2 Software de vizualizare multicriterială si de tip reţea SVMR**, pe baza unui sistem de filtrare multipla (http://radarrepository.uefiscdi.ro), care permite utilizatorilor accesul online pe baza unui cont cu credentiale unice.

Mecanismul de validare umana a presupus un proces continuu de lectura si incadrare a stirilor din baza de date in categoria „semnalelor slabe" sau a „non-semnalelor". Acest efort a revenit unei echipe formata din 20 de studenti masteranzi cu specializari diverse, recrutati pe baza abilitatilor de intelegere a textelor in limba engleza. Acestia din urma au fost organizati in echipe de cate doua persoane, schimbandu-si compozitia la fiecare flux de lucru saptamanal. Interactiunea s-a realizat online pe o platforma de evaluare colaborativa a stirilor tehnologice, denumita sugestiv, TAGy (**R7.1** 1 platforma operationala privind evaluarea colaborativă a ştirilor tehnologice, http://tagy.uefiscdi.ro/Account/Login.aspx), dezvoltata in cadrul proiectului.

Analiza automata presupune identificarea de *patternuri* cu ajutorul unor algoritmi specifici de clasificare de texte: LDA (Latent Dirichet Analysis) - pentru identificarea de pattern-uri independent de analiza umana si SVM (Support Vector Machine) - pentru predictie de semnale slabe si non-semnale pe baza exemplelor validate manual.

In vederea prelucrarii stirilor rezultate in urma evaluarii umane, inclusiv cu scopul furnizarii de materiale pregatitoare pentru analiza automata, a fost nevoie de construirea unei ontologii proprii pentru semnale slabe. Aceasta activitate a presupus crearea unor rețele semantice care să includă termenii specifici; cu alte

cuvinte a unor dictionare de termeni cu structura arborescenta (**R7.5 Un scenariu de analiza semantica),** în programul Tropes Zoom. Aceste dezvoltari permit incadrarea automata a stirilor cu acuratete pe taxonomii multiple, precum si vizualizarea lor multicriteriala.

Un alt modul dezvoltat in cadrul activitatii este acela de vizualizare a stirilor pe categorii semantice (**R7.6 Un modul de vizualizare a stirilor pe categorii semantice**, http://greuceanu.uefiscdi.ro/Login.aspx). Aceasta componenta permite si ierarhizarea unor seturi de stiri impartite pe categorii semantice si a fost folosit cu succes in cadrul unor exercitii de orientare strategica a Institutelor Nationale de Cercetare Dezvoltare (INCD). (Vezi anexa Platforma sondaje)

Pentru accesarea modulelor de software dezvoltate in cadrul proiectului, potentialii utilizatori pot solicita credentiale de acces la adresa info_radar@uefiscdi.ro. In cadrul anexelor acestui raport, pentru 2 dintre platforme(R7.1 si R7.2) au fost create conturi de acces (username si parola) in vederea vizualizarii datelor prezentate.

Platforma de management de cunoaștere la nivelul structurilor MECS responsabile în domeniul CDI şi al UEFISCDI a presupus realizarea, testarea şi operaţionalizarea urmatoarelor module:

I. **Dezvoltarea componentei de prevenire timpurie (*early warning*) a platformei de management de cunoaştere**

1.1. Colectarea textelor și analiza surselor

- Identificare unui set relevant de site-uri de stiri preponderent tehnologice din intreaga lume (minim 50)
- Crearea unei baze de stiri (repozitoriu) din aceste surse de stiri
- Analiza surselor

1.2. Validare umană

- Dezvoltare unei platforme tip *gaming* pentru evaluarea umana a stirilor
- Evaluarea umana a unui set de stiri folosind platforma dezvoltata (minim 10.000 stiri evaluate saptamanal)

II. **Implementarea componentei de analiză semantică din cadrul platformei de management de cunoaştere**

2.1. Dezvoltarea scenariului de analiza semantica

2.2. Dezvoltarea componentei de analiză automată a știrilor

În prezentul raport sunt descrise activitățile desfășurate pe fiecare componentă în parte.

I. Dezvoltarea componentei de prevenire timpurie (early warning) a platformei de management de cunoaştere

I.1. Colectarea textelor și analiza surselor

Un număr mare de știri (preponderent tehnologice) sunt preluate constant ca exemple/ resurse de învățare si prelucrate sub diferite formate pentru analize (analiza manuală - TAGy, analize semantice, analiza automată, analiza surselor). Știrile sunt preluate de pe o serie de platforme online anterior identificate dar și updatate pe parcursului proiectului.

De la începutul anului 2015, procesul de colectare a știrilor s-a automatizat prin colectarea lor din RSS feeder-uri. In perioada de raportare au fost colectate prin această metodă 193.431 de stiri. Pe tot parcursul proiectului s-au analizat informatiile si continutul stirilor astfel incat, pe de o parte unele dintre platforme au fost eliminate datorita faptului ca nu au generat suficiente resurse pentru identificarea tendintelor emergente iar pe de alta parte, au fost adaugate platforme noi, cu potential. Lista finala a resurselor online (platforme de stiri) este evidentiata in figura de mai jos, sau poate fi vizualizata accesand adresa: (http://192.168.10.21/RSSfeeder/) în *Fig.1.1 Lista finala platforme RSS feeder.*

Proiect cofinanţat din Fondul Social European, prin Programul Operaţional "Dezvoltarea Capacităţii Administrative", în perioada 2007-2013

Fig.1.1. Lista finala platforme RSS feeder

Situatie RSS feeder	
Nr crt	**Platforma**
1	www.firstpost.com
2	www.sciencedaily.com
3	www.technology.org
4	www.forbes.com/technology
5	techcrunch.com
6	www.huffingtonpost.com
7	www.entrepreneur.com
8	www.zdnet.com
9	phys.org
10	www.businessinsider.com/sai
11	www.venturebeat.com
12	www.gizmag.com
13	www.dailymail.co.uk/sciencetech
14	www.cleantechnica.com
15	www.techradar.com
16	www.salon.com
17	www.theguardian.com
18	www.firstpost.com/tech
19	www.biospace.com
20	www.scientificamerican.com
21	www.nanowerk.com
22	www.livescience.com
23	www.theverge.com/tech
24	www.foodnavigator.com
25	news.discovery.com
26	www.popsci.com
27	www.cbc.ca/news/technology
28	www.newscientist.com
29	www.iflscience.com
30	www.usatoday.com/tech/
31	www.biosciencetechnology.com
32	www.reuters.com/news/technology
33	www.azonano.com
34	www.space.com
35	www.electronicsweekly.com
36	www.extremetech.com
37	www.securityweek.com
38	www.naturalnews.com
39	www.science20.com
40	www.foxnews.com/tech
41	www.nanotech-now.com
42	blogs.sap.com/innovation
43	www.genengnews.com
44	www.independent.co.uk/life-style/gadgets-and-tech/
45	news.sciencemag.org
46	www.techinasia.com
47	www.environmentalleader.com
48	www.bbc.com/technology
49	newsoffice.mit.edu
50	campustechnology.com
51	www.greentechmedia.com
52	www.xconomy.com/channels
53	www.bbc.com/science_and_environment
54	www.3ders.org
55	www.foodnavigator-asia.com
56	www.npr.org/sections/technology
57	www.news.com.au/technology
58	spectrum.ieee.org
59	www.medgadget.com
60	timesofindia.indiatimes.com/home/science
61	www.3dprintingindustry.com
62	http://abcnews.go.com/Technology
63	www.perfscience.com

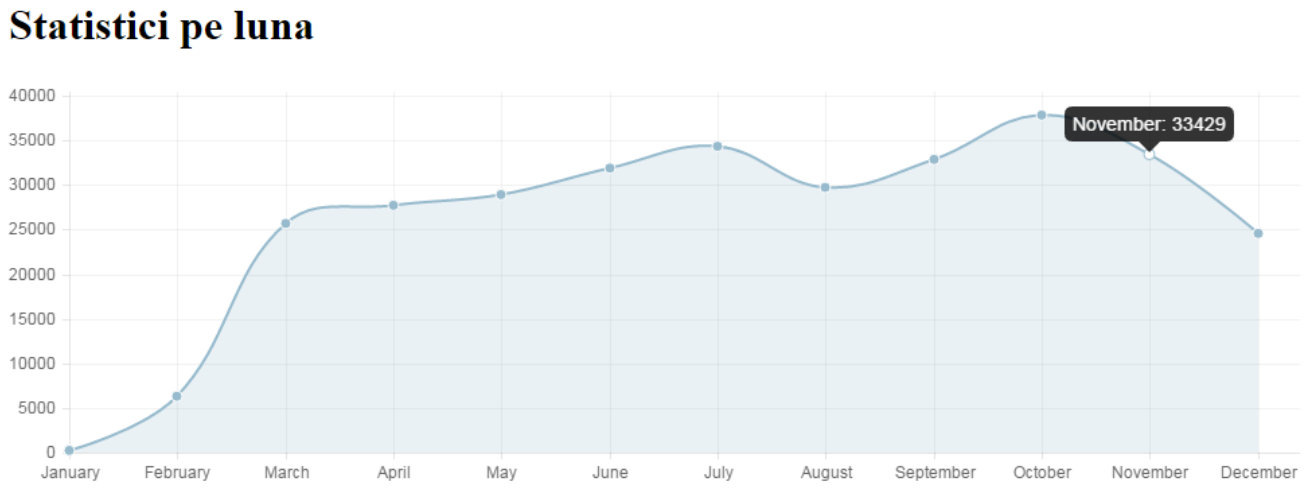
64	www.latimes.com/science
65	www.impactlab.net
66	www.sci-news.com
67	www.bbc.com/education
68	www.sciencealert.com
69	www.socialnewsdaily.com
70	www.goodnewsnetwork.org
71	www.edsurge.com
72	www.aerospace-technology.com
73	scitechdaily.com
74	www.washingtonpost.com/business/technology
75	www.healthcareitnews.com
76	uncovercalifornia.com/business/technology
77	www.technologyreview.com
78	www.biospectrumasia.com
79	robohub.org
80	smartcitiescouncil.com
81	www.azosensors.com
82	www.autonews.com
83	www.theverge.com/science
84	www.aerotechnews.com
85	www.the-scientist.com
86	www.recyclingtoday.com
87	searchenginewatch.com
88	www.sfgate.com/technology
89	www.technewsworld.com
90	http://gadgets.ndtv.com/news
91	profit.ndtv.com/news/banking-finance
92	www.abc.net.au
93	www.foodproductiondaily.com
94	www.in-pharmatechnologist.com
95	www.mnn.com
96	www.renewableenergyworld.com
97	www.psfk.com
98	www.inside3dprinting.com
99	www.roadtovr.com
100	www.autoweek.com
101	www.army-technology.com
102	www.eschoolnews.com
103	www.water-technology.net
104	www.neurosciencenews.com/neuroscience-topics
105	edition.cnn.com/TECH
106	www.theengineer.co.uk
107	www.theregister.co.uk/science
108	www.biopharma-reporter.com
109	www.business-standard.com/technology-news
110	www.npr.org/sections/research-news
111	www.photonics.com
112	www.cnbc.com/id/10000876
113	www.japantoday.com/category/technology
114	www.recyclinginternational.com
115	www.bbc.com/health
116	www.uncovermichigan.com/business/technology
117	www.designboom.com/technology
118	www.bbc.com/future
119	http://www.business-standard.com/category/finance
120	www.insidehpc.com/category/news-analysis/
121	ec.europa.eu/information_society
122	www.optics.org
123	www.technode.com
124	www.biotech-now.org
125	www.springwise.com
126	www.mainenewsonline.com/business/technology
127	utsandiego.com/headlines/business/technology
128	biopharma-asia.com
129	www.todayonline.com/tech

130	www.augmentedrealitytrends.com
131	www.japantimes.co.jp/tech
132	www.singularityhub.com
133	www.nordiclifescience.org
134	www.nanomagazine.co.uk/category&id=172&Itemid
135	www.bionews.org.uk
136	www.wsj.com/news/technology
137	defensetech.org
138	www.waste-management-world.com/recycling.html
139	www.pbs.org/wgbh/nova/next/
140	www.watertechonline.com
141	www.biologynews.net/
142	www.euronews.com/sci-tech/
143	www.enn.com
144	www.technabob.com
145	www.postscapes.com/internet-of-things-news
146	defense-update.com
147	www.chinatechnews.com
148	www.nsf.gov
149	www.shapeways.com/blog
150	www.nasa.gov
151	www.asmarterplanet.com
152	www.itechfuture.com
153	www.european-biotechnology-news.com
154	www.economist.com/sections/science-technology
155	www.gigaom.com/channel/science-energy
156	www.theaustralian.com.au/business/technology
157	www.economist.com/sections/international
158	www.foodtechnology.co.nz
159	www.watertechonline.com/topics/14555-municipal
160	www.wfs.org/category/user-interest-tags/scitech
161	www.watertechonline.com/topics/6122-Drinking-Wa
162	news.stanford.edu/news/socsci/
163	www.mininginnovationnews.com/category/new-tech
164	www.moreinspiration.com
165	www.watertechonline.com/topics/6126-environmen
166	www.watertechonline.com/topics/14556-Industrial
167	www.alternative-energy-news.info
168	www.neurogadget.com
169	www.rand.org/topics/terrorism-and-homeland-secur
170	www.forumforthefuture.org
171	www.collective-evolution.com/category/sci-tech
172	www.frontlinedesk.com/science-and-fiction
173	www.itwire.com/itwire-rss/science-news
174	www.voicechronicle.com/category/tech-and-science
175	www.telecomstechnews.com
176	www.rand.org/topics/law-and-business
177	www.techthefuture.com
178	www.rand.org/topics/science-and-technology
179	www.rand.org/topics/education-and-the-arts
180	www.iotnewsnetwork.com
181	www.rand.org/topics/health-and-health-care
182	www.clickgreen.org.uk/rss/research
183	www.economist.com/topics/life-sciences
184	www.frontlinedesk.com/technology
185	phidgets.wordpress.com
186	www.rand.org/topics/cyber-warfare
187	www.theguardian.com/society/social-trends
188	www.augmented.org/blog
189	www.news-medical.net/?tag=/Biosensor
190	www.rand.org/topics/children-and-families
191	www.usnews.com/topics/subjects/biology
192	www.techanalyst.co
193	www.greenwisebusiness.co.uk

194	www.waste-management-world.com/recycling.html
195	www.eurekalert.org/bysubject/mathematics.php
196	www.eurekalert.org/bysubject/education.php
197	www.eurekalert.org/bysubject/oceanography.php
198	www.clickgreen.org.uk/rss/research
199	www.watertechonline.com/topics/6122-Drinking-Water
200	www.cnbc.com/id/44877279
201	www.foodtechnology.co.nz
202	www.gigaom.com/channel/science-energy
203	www.treehugger.com/business
204	www.watertechonline.com/topics/14556-Industrial
205	www.treehugger.com/technology
206	www.collective-evolution.com/category/sci-tech
207	www.nasa.gov
208	physicsworld.com
209	www.european-biotechnology-news.com
210	www.treehugger.com/energy
211	www.techthefuture.com
212	humanoids.io
213	www.iotnewsnetwork.com
214	www.mininginnovationnews.com/category/new-technology
215	www.nytimes.com/business/smallbusiness
216	www.itwire.com/itwire-rss/science-news
217	www.economist.com/topics/life-sciences

217	www.economist.com/topics/life-sciences
218	www.forbes.com/green-tech
219	www.alternative-energy-news.info
220	www.rand.org/topics/terrorism-and-homeland-security
221	www.theguardian.com/society/social-trends
222	www.news-medical.net/?tag=/Biosensor
223	www.economist.com/topics/space-technology
224	www.rand.org/topics/law-and-business
225	www.voicechronicle.com/category/tech-and-science
226	www.rand.org/topics/science-and-technology
227	www.frontlinedesk.com/science-and-fiction
228	www.rand.org/topics/health-and-health-care
229	www.rand.org/topics/education-and-the-arts
230	www.usnews.com/topics/subjects/biology
231	www.rand.org/topics/cyber-warfare
232	www.augmented.org/blog
233	www.greenwisebusiness.co.uk
234	phidgets.wordpress.com
235	www.quantamagazine.org/category/biology-2
236	www.frontlinedesk.com/technology
237	www.rand.org/topics/children-and-families
238	www.quantamagazine.org/category/computer-physics-2
239	www.quantamagazine.org/category/computer-science-2
240	www.techanalyst.co
241	www.emdt.co.uk

Fig. 1.2. Distributia lunara a colectarii de stiri, 2015



Știrile preluate prin RSS feeder au fost transferate în platforma Repository (această platformă cuprinde toate știrile colectate de-a lungul timpului, inclusiv prin RSS feeder; știrile sunt identificate prin denumirea care conține un cod txt, numele platformei de care aparține știrea, în format txt dar și Excel).

Pentru o selecție mai eficientă a ştirilor care sunt supuse validării umane prin platforma Tagy, s-a realizat o ancoră între cele două platforme (Tagy şi Repository), astfel automatizându-se mai multe procese care stăteau la baza activităților asociate celor două instrumente (*vezi* descrierea acestora două în ***Anexa 1_Repository_Tagy***).

I.1.1. Analiza surselor

Analiza și clasificare surse

Pornind de la sursele amintite mai sus s-a realizat o analiză, atât din punct de vedere cantitativ (numărul de știri preluate din RSS feeder și introduse în platforma Tagy, data ultimei extrageri, numarul de intrări in Tagy) dar și calitativ (procentul de semnale slabe pentru fiecare platforma a platformelor de știri (241 de platforme). Rezultatele acestei analize sunt utile pentru clasificarea platformelor de știri în funcție de relevanța informațiilor furnizate (cu cât numărul de SS este mai mare raportat la numărul de știri votate, cu atât este mai relevantă platforma în sine). O previzualizare a clasificării platformelor se poate observa în *figura 2* de mai jos.

Fig. 2. Captura clasificare platforme

Platforma	Data ultimei extrageri din RSS	Nr. Stiri RSS (txt) 02.07	Nr intrari TAGy 15.07	Nr. SS(TAGy)	Nr. NS(TAGY)	% SS TAGy
http://www.3ders.org/	30-Jun-15	528	422	28	394	6.64%
http://www.news-medical.net/?tag=/Biosensor	24-Jun-15	10	7	3	4	42.86%
http://nanomagazine.co.uk/index.php?option=com_sectionex&view=category&id=172&Itemid=158	30-Jun-15	100	58	19	39	32.76%
http://www.moreinspiration.com/article/6144/control-windsreen-wipers-with-your-eyes	22-Jun-15	33	10	3	7	30.00%
http://www.alternative-energy-news.info/	13-Apr-2015	27	4	1	3	25.00%
http://3dprintingindustry.com/	30-Jun-15	464	266	11	255	4.14%
http://www.abc.net.au/science/news/	30-Jun-15	304	236	2	234	0.85%
http://www.aerotechnews.com/news/	30-Jun-15	318	283	4	279	1.41%
http://neurogadget.com/	24-Jun-15	26	21	4	17	19.05%
http://www.augmentedrealitytrends.com/	30-Jun-15	110	138	2	136	1.45%
http://www.azonano.com/	30-Jun-15	925	493	68	425	13.79%
http://www.azosensors.com/	30-Jun-15	338	424	34	390	8.02%

Pentru a scoate în evidență câteva rezultate ale analizei mai sus menționate se poate observa numarul de platforme de știri care au mai mult de 50/100/200/300/500/1000 de stiri precum si numarul de platforme care depășesc procentul de 2% semnale slabe conform procesului de vot din Tagy.

Stiri in Tagy	Numar platforme
>50	148
>100	120
>200	82
>300	58
>500	40
>1000	19

% SS	Numar platforme (cu peste 50 de stiri in Tagy)
>10%	20
>5%	44
>3%	59
>2%	74

% SS	Numar platforme(cu peste 100 de stiri in Tagy)
>10%	18
>5%	39
>3%	54
>2%	66

Analiza similitudinii

O parte din textele colectate sunt analizate cu software-ul de plagiat în vederea identificării de surse adiționale pentru fiecare știre în parte. Sunt astfel generate rapoarte ce prezintă procente ale similarității știrilor cu texte din alte surse de pe internet.

Pentru completarea analizei a fost dezvoltat intern un soft pentru preluarea datei știrilor din toate sursele identificate.

În perioada raportată a fost realizată o analiză a surselor unui eșantion de știri din Repository (din aproximativ 60 de platforme) cu ajutorul software-ului Plagiarism Detector și al software-ului dezvoltat intern de preluare automată a știrilor și transformarea rezultatelor în format excel. Astfel, a rezultat documentul Anexa 2.Rezultate surse aditionale stiri similitudine . Acesta prezintă aproximativ 700 de știri pentru care au fost identificate câte alte 10 surse pe care se repetă fiecare știre cu minim 50% procent de similaritate. De asemenea, data știrii din fiecare sursă a fost extrasă și este vizibilă în excelul cu rezultate. Din analiză au rezultat și o serie de site-uri de tip agregator, prezentate în figura de mai jos:

Fig. 3. Agregatoare identificate în analiza surselor

Nr crt	Agregatoare
1	http://www.ideaslaboratory.com/ (AGG)
2	http://able2know.org/topic/226001-132 (AGG)
3	http://campustechnology.com/articles/2014/11/20/stanford-researchers-create-computer-vision-algorithm-for-describing-visual-scenes.aspx (AGG)
4	http://cordis.europa.eu/news/rcn/122134_en.pdf (AGG)
5	http://euroiphonenews.com/ (AGG)
6	http://jqi.umd.edu/news/entropy-nations (AGG)
7	http://lightspeedindia.wordpress.com/ (AGG)
8	http://nanocomputer.com/ (AGG)
9	http://neuronclub.org/2014/01/ (AGG)
10	http://newenergyandfuel.com/ (AGG)
11	http://porqueno.tumblr.com/ (AGG)
12	http://potentiamed.com/category/darpa/ (AGG)
13	http://programacion2z.wordpress.com/ (AGG)
14	http://redmondmag.com/home.aspx?Page=5 (AGG)
15	http://scienceofsingularity.com/ (AGG)
16	http://semimd.com/blog/tag/soitec/ (AGG)
17	http://sheerwind.com/media/releases (AGG)
18	http://southendaquarist.weebly.com/ (AGG)
19	http://vivoforums.com/news/ (AGG)
20	http://wiche.edu/state-highlights/idaho (AGG)
21	http://www.aacn.org/wd/Cetests/media/A142301.pdf (AGG)
22	http://www.ahpanet.com/?page=LatestNews (AGG)
23	http://www.anl.gov/news-room (AGG)
24	http://www.auburnsentinel.com/ (AGG)
25	http://www.bioquicknews.com/node/1505 (AGG)
26	http://www.brijj.com/vk-gupta (AGG)
27	http://www.ehrexperts.us/ (AGG)
28	http://www.futuretimeline.net/blog.htm (AGG)
29	http://www.ox.ac.uk/news/science-blog (AGG)
30	http://www.uc.edu/news/NR.aspx?id=20860 (AGG)
31	http://www.umdrightnow.umd.edu/ (AGG)
32	http://www.veooz.com/news/xHWsVg5.html (AGG)
33	http://www.wiche.edu/sara (AGG)

Acest pas va ajuta la următoarele analize de suprapunere și comparare a surselor în vederea:

- Realizării unui top al surselor de știri;
- Indentificării de noi radare;
- Identificării sursei primare a știrilor.

Tot în perioada raportată au fost realizate câteva analize cu software-ul de plagiat *Plagiarism Detector* pe un eșantion de știri validate ca semnale slabe în platforma TAGy. 1309 stiri au fost trecute prin software-ul de detectare a similiaritatii, astfel au fost generate 1309 rapoarte. O parte din rapoartele rezultate pot fi consultate în ***Anexa 3. Rapoarte PD SS*** (125 rapoarte).

Din cele 1309 rapoarte doar pentru 781 de ştiri au fost identificate surse cu procent de minim 50%.

De asemenea au fost identificate şi agregatoare (o parte dintre acestea pot fi vizualizate mai jos). Aceste surse nu au fost adăugate in RSS.

Proiect cofinanţat din Fondul Social European, prin Programul Operaţional "Dezvoltarea Capacităţii Administrative", în perioada 2007-2013

Fig.4 Agregatoare

1	http://andreswhy.blogspot.com/ (AGG)
2	http://arrgt13.rssing.com/chan-7597928/all_p5.html (AGG)
3	http://berkeley914.rssing.com/chan-28666368/all_p1.html (AGG)
4	http://biodips.com/a/00000000b2d63fb9 (AGG)
5	http://cambridge597.rssing.com/chan-11168856/all_p2.html (AGG)
6	http://cancerlive.net/ (AGG)
7	http://cars.psghost.net/tag/solar/ (AGG)
8	http://cavendish81.rssing.com/chan-11168859/all_p3.html (AGG)
9	http://chayroot57.rssing.com/chan-27480809/all_p1.html (AGG)
10	http://chrysobalanaceae57.rssing.com/chan-27523191/all_p2.html (AGG)
11	http://corsy14.rssing.com/chan-8100371/all_p3.html (AGG)
12	http://coulterneb14.rssing.com/chan-8101289/all_p6.html (AGG)
13	http://daily-nano-news.blogspot.com/2015/05/microcombing-creates-stronger-more.html (AGG)
14	http://dawdlers4.rssing.com/chan-3600489/all_p6.html (AGG)
15	http://dcircuits.com/?paged=3&m=201505 (AGG)
16	http://discountbook.in/?paged=2 (AGG)
17	http://drumbledore55.rssing.com/chan-24107249/all_p1.html (AGG)
18	http://drunkschmuck.tumblr.com/page/70 (AGG)
19	http://emvco.us/?p=54 (AGG)
20	http://engineering.rice.edu/biomimetic/ (AGG)
21	http://forum.boinaslava.net/showthread.php?13662-%CA%EE%F1%EC%E8%F7%E5%F1%EA%E8-%ED%EE%E2%E8%ED%E8/page
22	http://forums.spacebattles.com/threads/ibm-sets-new-tape-storage-record.335843/ (AGG)
23	http://genesisnanotech.com/solar-cell/ (AGG)
24	http://geneticenter.com/2013/10/ (AGG)
25	http://godwincaruana.me/tag/research/ (AGG)
26	http://helmet511.rssing.com/chan-33977129/all_p1.html (AGG)
27	http://hulotheism.rssing.com/chan-1613869/all_p274.html (AGG)
28	http://ieeeitesm.org/news/ (AGG)
29	http://industry1090.rssing.com/chan-4098869/all_p10.html (AGG)
30	http://jaunders55.rssing.com/chan-24149330/all_p1.html (AGG)
31	http://lidar82.rssing.com/chan-28258599/all_p1.html (AGG)
32	http://light783.rssing.com/chan-14382296/all_p2.html (AGG)
33	http://ma.webradar.me/portal/40012187 (AGG)

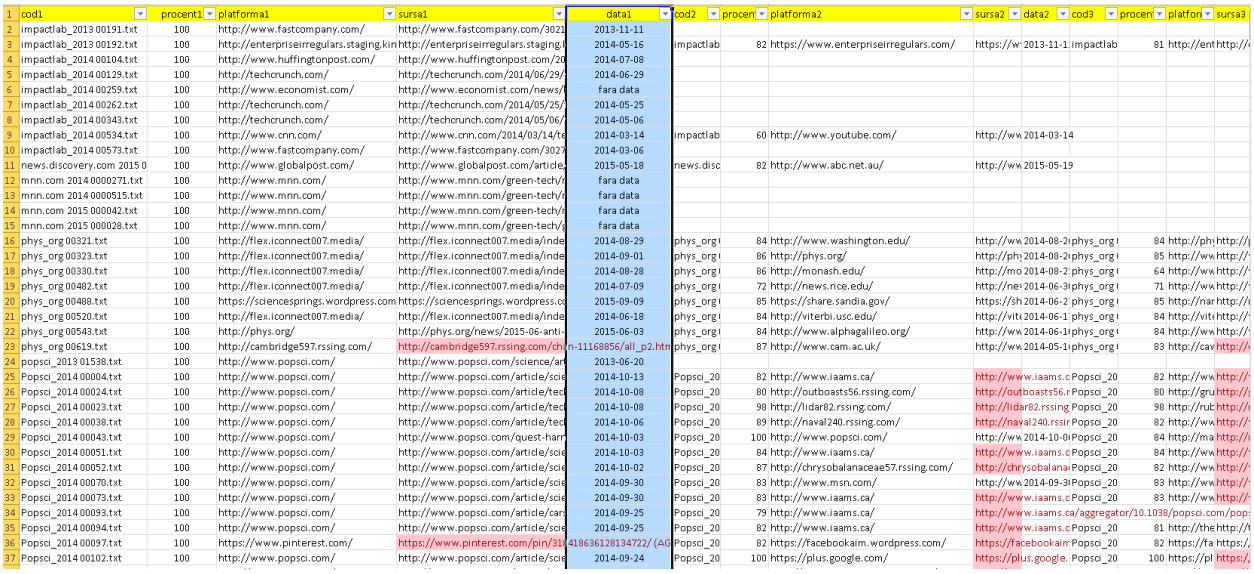
34	http://machines158.rssing.com/chan-23604288/all_p2.html (AGG)
35	http://mashable.com/category/dna/ (AGG)
36	http://mispleaded56.rssing.com/chan-24596183/all_p2.html (AGG)
37	http://mixedcrunch.blogspot.com/ (AGG)
38	http://nanobrainimplant.com/ (AGG)
39	http://naval240.rssing.com/chan-10893106/all_p2.html (AGG)
40	http://news.dxy.cn/bbs/thread/30820918 (AGG)
41	http://news.wisc.edu/releases/19049 (AGG)
42	http://noncustodial56.rssing.com/chan-24737094/all_p1.html (AGG)
43	http://otididae56.rssing.com/chan-24957105/all_p1.html (AGG)
44	http://outboasts56.rssing.com/chan-24937888/all_p2.html (AGG)
45	http://ow.ly/LBgtj (AGG)
46	http://pests394.rssing.com/chan-28610491/all_p1.html (AGG)
47	http://plug-ugly3.rssing.com/chan-3371028/all_p82.html (AGG)
48	http://plug-ugly3.rssing.com/chan-3371028/all_p85.html (AGG)
49	http://prepayments12.rssing.com/chan-7399440/all_p3.html (AGG)
50	http://princeton675.rssing.com/chan-30037752/all_p3.html (AGG)
51	http://processing284.rssing.com/chan-19070398/latest.php (AGG)
52	http://projectavalon.net/forum4/showthread.php?62082-Technological-advances-that-will-directly-affect-you-in-the-next-2-yea
53	http://pseudofeminine56.rssing.com/chan-25453727/all_p2.html (AGG)
54	http://reawakes56.rssing.com/chan-25545354/all_p2.html (AGG)
55	http://reddylab.org/news/ (AGG)
56	http://research.uconn.edu/2014/09/ (AGG)
57	http://roscosmos1.rssing.com/chan-19077869/all_p1.html (AGG)
58	http://rustiness56.rssing.com/chan-25708429/all_p4.html (AGG)
59	http://safety1713.rssing.com/chan-8246759/all_p1.html (AGG)
60	http://sarahazablog.tumblr.com/ (AGG)
61	http://savings695.rssing.com/chan-7053140/all_p49.html (AGG)
62	http://scenographical53.rssing.com/chan-23736376/all_p1.html (AGG)
63	http://schen.ucsd.edu/lab/news/Kurzweil.pdf (AGG)
64	http://semidivisively56.rssing.com/chan-25857449/all_p1.html (AGG)
65	http://semimd.com/blog/tag/imec/ (AGG)
66	http://sewers71.rssing.com/chan-34134844/all_p1.html (AGG)

http://www.bioquicknews.com/node/2391 (AGG)
http://www.bioquicknews.com/node/2473 (AGG)
http://www.bioquicknews.com/node/2530 (AGG)
http://www.bioquicknews.com/node/2532 (AGG)
http://www.bioquicknews.com/node/2553 (AGG)
http://www.bluesoleil.com/Life/281.html (AGG)
http://www.caltech.edu/tags/energy (AGG)
http://www.cs.berkeley.edu/ (AGG)
http://www.eduloc.com/hot_news/773.html (AGG)
http://www.elektrosmog.com/english-1/ (AGG)
http://www.frogheart.ca/?tag=silicene (AGG)
http://www.hearingaidforums.com/showthread.php?16385-Cochlear-Implants-with-No-Exterior-Hardware-Its-Here (AGG)
http://www.helpcash.gq/microloans.html (AGG)
http://www.iaams.ca/aggregator/10.1038/magazine?page=45 (AGG)
http://www.notey.com/text/blogs/typing (AGG)
http://www.taodocs.com/p-10117841.html (AGG)
http://www.therxforum.com/showthread.php?t=725297&p=11042860 (AGG)
http://www.uc.edu/news/NR.aspx?id=20860 (AGG)
http://www.waseda.jp/top/en-news/25025 (AGG)
http://www.wfs.org/aggregator/sources/4?page=7 (AGG)
http://www.xradia.com/press-releases/ (AGG)
https://facebookaim.wordpress.com/ (AGG)
https://hisaurav.wordpress.com/ (AGG)
https://medicupdate.wordpress.com/ (AGG)
https://plus.google.com/events/c1lnj82cdksanhko39e1g1cg1fk (AGG)
https://www.broadinstitute.org/files/news/media-kit/2015/CRISPR-FengZhang-presskit.pdf (AGG)
https://www.facebook.com/hot877/posts/10152910257361220:0 (AGG)
https://www.facebook.com/pages/Inventers-Hub/282412521925675?fref=photo (AGG)
https://www.facebook.com/pages/Wonder-Of-The-Science/1511872849028922 (AGG)
https://www.facebook.com/permalink.php?id=472130596221947&story_fbid=628078567293815 (AGG)
https://www.hzdr.de/db/Cms?pOid=44032 (AGG)
https://www.pinterest.com/autonomousnow/autonomous-vehicles-news/ (AGG)
https://zedie.wordpress.com/tag/iphone/ (AGG)

Toate rapoartele au fost transpuse într-un excel care conţine informaţia de care aveam nevoie pentru analiză.

O parte din documentul excel poate fi vizualizat în imaginile de mai jos, iar documentul integral poate fi consultat in Anexa 2.Rezultate surse aditionale stiri similitudine.

Fig.4 Vizualizare identificare surse stiri



	cod1	procent1	platforma1	sursa1	data1	cod2	procen	platforma2	sursa2	data2	cod3	procen	platfor	sursa3
2	impactlab_2013 00191.txt	100	http://www.fastcompany.com/	http://www.fastcompany.com/3021	2013-11-11									
3	impactlab_2013 00192.txt	100	http://enterpriseirregulars.staging.kin	http://enterpriseirregulars.staging.k	2014-05-16	impactlab	82	https://www.enterpriseirregulars.com/	https://w	2013-11-1	impactlab	81	http://ent	http://e
4	impactlab_2014 00104.txt	100	http://www.huffingtonpost.com/	http://www.huffingtonpost.com/20	2014-07-08									
5	impactlab_2014 00129.txt	100	http://techcrunch.com/	http://techcrunch.com/2014/06/29/	2014-06-29									
6	impactlab_2014 00259.txt	100	http://www.economist.com/	http://www.economist.com/news/t	fara data									
7	impactlab_2014 00262.txt	100	http://techcrunch.com/	http://techcrunch.com/2014/05/25/	2014-05-25									
8	impactlab_2014 00343.txt	100	http://techcrunch.com/	http://techcrunch.com/2014/05/06/	2014-05-06									
9	impactlab_2014 00534.txt	100	http://www.cnn.com/	http://www.cnn.com/2014/03/14/te	2014-03-14	impactlab	60	http://www.youtube.com/	http://ww	2014-03-14				
10	impactlab_2014 00573.txt	100	http://www.fastcompany.com/	http://www.fastcompany.com/3027	2014-03-06									
11	news.discovery.com 2015 0	100	http://www.globalpost.com/	http://www.globalpost.com/article_	2015-05-18	news.disc	82	http://www.abc.net.au/	http://ww	2015-05-19				
12	mnn.com 2014 0000271.txt	100	http://www.mnn.com/	http://www.mnn.com/green-tech/t	fara data									
13	mnn.com 2014 0000515.txt	100	http://www.mnn.com/	http://www.mnn.com/green-tech/t	fara data									
14	mnn.com 2015 000042.txt	100	http://www.mnn.com/	http://www.mnn.com/green-tech/t	fara data									
15	mnn.com 2015 000028.txt	100	http://www.mnn.com/	http://www.mnn.com/green-tech/g	fara data									
16	phys_org 00321.txt	100	http://flex.iconnect007.media/	http://flex.iconnect007.media/inde	2014-08-29	phys_org	84	http://www.washington.edu/	http://ww	2014-08-2	phys_org	84	http://ph	http://p
17	phys_org 00323.txt	100	http://flex.iconnect007.media/	http://flex.iconnect007.media/inde	2014-09-01	phys_org	86	http://phys.org/	http://ph	2014-08-2	phys_org	85	http://ww	http://p
18	phys_org 00330.txt	100	http://flex.iconnect007.media/	http://flex.iconnect007.media/inde	2014-08-28	phys_org	86	http://monash.edu/	http://mo	2014-08-2	phys_org	64	http://ww	http://p
19	phys_org 00482.txt	100	http://flex.iconnect007.media/	http://flex.iconnect007.media/inde	2014-07-09	phys_org	72	http://news.rice.edu/	http://ne	2014-06-3	phys_org	71	http://ww	http://p
20	phys_org 00488.txt	100	https://sciencesprings.wordpress.com	https://sciencesprings.wordpress.co	2015-09-09	phys_org	85	https://share.sandia.gov/	https://sh	2014-06-2	phys_org	85	http://nar	http://s
21	phys_org 00520.txt	100	http://flex.iconnect007.media/	http://flex.iconnect007.media/inde	2014-06-18	phys_org	84	http://viterbi.usc.edu/	http://vit	2014-06-1	phys_org	84	http://vit	http://v
22	phys_org 00543.txt	100	http://phys.org/	http://phys.org/news/2015-06-anti-	2015-06-03	phys_org	84	http://www.alphagalileo.org/	http://ww	2014-06-1	phys_org	84	http://ww	http://a
23	phys_org 00619.txt	100	http://cambridge597.rssing.com/	http://cambridge597.rssing.com/ch	n-11168656/all_p2.htm	phys_org	87	http://www.cam.ac.uk/	http://ww	2014-05-1	phys_org	83	http://cav	http://c
24	popsci_2013 01538.txt	100	http://www.popsci.com/	http://www.popsci.com/science/ar	2013-06-20									
25	Popsci_2014 00004.txt	100	http://www.popsci.com/	http://www.popsci.com/article/scie	2014-10-13	Popsci_20	82	http://www.iaams.ca/	http://ww	2014-10-1	Popsci_20	82	http://ww	http://i
26	Popsci_2014 00024.txt	100	http://www.popsci.com/	http://www.popsci.com/article/tech	2014-10-08	Popsci_20	80	http://outboasts56.rssing.com/	http://ou	2014-10-0	Popsci_20	80	http://gru	http://i
27	Popsci_2014 00023.txt	100	http://www.popsci.com/	http://www.popsci.com/article/tech	2014-10-08	Popsci_20	98	http://lidar82.rssing.com/	http://lid	2014-10-0	Popsci_20	98	http://rub	http://i
28	Popsci_2014 00038.txt	100	http://www.popsci.com/	http://www.popsci.com/article/tech	2014-10-06	Popsci_20	89	http://naval240.rssing.com/	http://na	2014-10-0	Popsci_20	82	http://ww	http://i
29	Popsci_2014 00043.txt	100	http://www.popsci.com/	http://www.popsci.com/quest-harr	2014-10-03	Popsci_20	100	http://www.popsci.com/	http://ww	2014-10-0	Popsci_20	84	http://ma	http://i
30	Popsci_2014 00051.txt	100	http://www.popsci.com/	http://www.popsci.com/article/scie	2014-10-03	Popsci_20	84	http://www.iaams.ca/	http://ww	2014-10-0	Popsci_20	84	http://ww	http://i
31	Popsci_2014 00052.txt	100	http://www.popsci.com/	http://www.popsci.com/article/scie	2014-10-02	Popsci_20	87	http://chrysobalanaceae57.rssing.com/	http://ch	rysobalanac	Popsci_20	84	http://ww	http://i
32	Popsci_2014 00070.txt	100	http://www.popsci.com/	http://www.popsci.com/article/scie	2014-09-30	Popsci_20	83	http://www.msn.com/	http://ww	2014-09-3	Popsci_20	83	http://ww	http://i
33	Popsci_2014 00073.txt	100	http://www.popsci.com/	http://www.popsci.com/article/scie	2014-09-30	Popsci_20	83	http://www.iaams.ca/	http://ww	.iaams.c	Popsci_20	83	http://ww	http://i
34	Popsci_2014 00093.txt	100	http://www.popsci.com/	http://www.popsci.com/article/cars	2014-09-25	Popsci_20	79	http://www.iaams.ca/	http://ww	.iaams.ca/	aggregator/10.1038/popsci.com/pop			
35	Popsci_2014 00094.txt	100	http://www.popsci.com/	http://www.popsci.com/article/scie	2014-09-25	Popsci_20	82	http://www.iaams.ca/	http://ww	.iaams.c	Popsci_20	81	http://the	http://t
36	Popsci_2014 00097.txt	100	https://www.pinterest.com/	https://www.pinterest.com/pin/31	418636128134722/ (AG	Popsci_20	82	https://facebookaim.wordpress.com/	https://fa	cebookaim	Popsci_20	82	https://fa	https://a
37	Popsci_2014 00102.txt	100	http://www.popsci.com/	http://www.popsci.com/article/scie	2014-09-24	Popsci_20	100	https://plus.google.com/	https://pl	us.google.	Popsci_20	100	https://pl	https://p

	cod1	procent1	platforma1	sursa1	data1	cod2	procen	platforma2	sursa2	data2	cod3	procen	platfor	sursa3
743	phys_org 00621.txt	78	http://www.uh.edu/	http://www.uh.edu/news-events/s	2014-05-15	phys_org	73	http://www.cailiaoren.com/	http://ww	fara data				
744	R_phys.org 2015 00003086.t	78	http://www.nytimes.com/	http://www.nytimes.com/aponline/	2015-04-14	R_phys.or	77	http://www.heraldcourier.com/	http://ww	2014-05-2	R_phys.or	77	http://ww	http://i
745	R_www.forbes.com_techn	78	http://news.newsdirectory1.com/	http://news.newsdirectory1.com/cs	2015-09-10									
746	impactlab_2013 00186.txt	77	http://www.impactlab.net/	http://www.impactlab.net/2013/11/	2013-11-15	impactlab	73	http://www.theverge.com/	http://ww	2013-11-13				
747	impactlab_2014 00052.txt	77	http://blogs.wsj.com/	http://blogs.wsj.com/indiarealtime/	2014-07-24	impactlab	65	http://www.nature.com/	http://ww	2012-12-0	impactlab	65	http://tec	http://t
748	impactlab_2014 00469.txt	77	http://mashable.com/	http://mashable.com/2014/04/04/ar	2014-04-04									
749	Nature 04407.txt	77	http://www.nature.com/	http://www.nature.com/news/whe	2013-02-13									
750	R_www.goodnewsnetwork	77	http://www.goodnewsnetwork.org/	http://www.goodnewsnetwork.org	2015-05-12									
751	impactlab_2014 00001.txt	76	http://singularityhub.com/	http://singularityhub.com/2014/08/	2014-08-05									
752	impactlab_2014 00119.txt	76	http://singularityhub.com/	http://singularityhub.com/2014/06/	2014-06-26									
753	news.discovery.com 2015 0	76	http://arstechnica.com/	http://arstechnica.com/tech-policy/	2015-05-19									
754	Popsci_2014 00014.txt	76	http://www.tumblr.com/	http://www.tumblr.com/tagged/bri	fara data	Popsci_20	76	http://www.popsci.com/	http://ww	2014-10-1	Popsci_20	75	http://ww	http://i
755	R_www.inside3dprinting.c	76	http://inside3dprinting.com/	http://inside3dprinting.com/neri-o	2015-05-15	R_www.ir	64	http://www.gizmag.com/	http://ww	2015-05-1	R_www.ir	63	http://ww	http://i
756	newsoffice.mit.edu 2015 0	75	http://nanobrainimplant.com/	http://nanobrainimplant.com/ (AGG)		newsoffic	75	http://nanobrainimplant.com/	http://nar	1999-08-20				
757	popsci_2013 00013.txt	75	http://www.popsci.com/	http://www.popsci.com/article/gad	fara data									
758	popsci_2013 00015.txt	75	http://machines158.rssing.com/	http://machines158.rssing.com/chan-23604288/all_p2.html		popsci_20	64	https://www.youtube.com/	https://w	fara data				
759	R_www.azonano.com 2015	75	http://www.angstronmaterials.com/	http://www.angstronmaterials.com/	2015-06-04	R_www.a	74	http://www.angstronmaterials.com/	http://ww	2015-06-0	R_www.a	74	https://w	https://a
760	neurosciencenews.com 201	74	http://www.eurekalert.org/	http://www.eurekalert.org/pub_rel	fara data	neuroscie	73	http://phys.org/	http://ph	2015-01-15				
761	popsci_2013 00155.txt	74	http://www.popsci.com/	http://www.popsci.com/article/tech	2013-10-24	popsci_20	68	http://scinewsnetwork.tumblr.com/	http://sci	fara data	popsci_20	67	https://ht	https://p
762	phys_org 00005.txt	73	http://semimd.com/	http://semimd.com/blog/tag/imec/ (AGG)										
763	popsci_2013 00091.txt	73	http://www.popsci.com/	http://www.popsci.com/technology	2013-09-10									
764	neurogadget.com 2015 000	70	http://www.technologyreview.com/	http://www.technologyreview.com/	2015-01-14									
765	phys_org 00264.txt	70	http://insideevs.com/	http://insideevs.com/rapid-charging	fara data									
766	popsci_2013 00067.txt	70	http://sigalontech.soup.io/	http://sigalontech.soup.io/tag/archi	fara data	popsci_20	70	http://www.popsci.com/	http://ww	2013-09-2	popsci_20	70	http://cac	http://e
767	neurosciencenews.com 201	69	http://www.eurekalert.org/	http://www.eurekalert.org/pub_rel	2015-02-04									
768	impactlab_2013 00188.txt	68	http://wtfrly.com/	http://wtfrly.com/2013/11/16/3d-pr	2013-11-16	impactlab	54	http://textile-future.com/	http://tex	fara data				
769	newsoffice 00471.txt	67	http://www.hearingaidforums.com/	http://www.hearingaidforums.com/showthread.php?16385-Cochlear-Implants-with-No-Exterior-Hardware-Its-Here (AGG)										
770	neurosciencenews.com 201	66	http://www.pratt.duke.edu/	http://www.pratt.duke.edu/news/f	2015-01-13									
771	popsci_2013 02418.txt	66	http://ep.yimg.com/	http://ep.yimg.com/ty/cdn/reality	2012-05-18									
772	newsoffice 00597.txt	64	http://scienceblogs.com/	http://scienceblogs.com/startswith	2011-12-05									
773	phys_org 00036.txt	64	http://www.israel21c.org/	http://www.israel21c.org/news/wir	2014-12-07									
774	neurosciencenews.com 201	62	http://www.dailymail.co.uk/	http://www.dailymail.co.uk/science	2015-02-19									
775	R_www.nanotech-now.com	60	http://www.news.ucsb.edu/	http://www.news.ucsb.edu/2015/0	2015-05-11	R_www.n	57	http://www.supercomputingonline.com/	http://ww	fara data				
776	impactlab_2014 00050.txt	58	http://thediagonal.com/	http://thediagonal.com/ideasoup/ (AGG)										
777	phys.org 2015 000056.txt	58	http://www.iranreview.org/	http://www.iranreview.org/content	2015-02-04									
778	mnn.com 2014 000025.txt	56	http://msutoday.msu.edu/	http://msutoday.msu.edu/news/20	2014-08-19	mnn.com	55	http://www.futurity.org/	http://ww	2014-08-20				
779	newsoffice 00392.txt	56	http://www.kurzweilai.net/	http://www.kurzweilai.net/replacir	2014-04-10									
780	popsci_2013 03178.txt	56	http://roscosmos1.rssing.com/	http://roscosmos1.rssing.com/chan-19077869/all_p1.html (AGG)										
781	impactlab_2014 00403.txt	52	http://www.youtube.com/	http://www.youtube.com/watch?v=	2014-04-10									

Tot pentru o clasificare a surselor, în platforma de evaluarea colaborativă a ştirilor (Tagy) am creat o zonă destinată votului platformelor pentru fiecare flux. Astfel, fiecare evaluator uman trebuia să acorde un scor platformelor de pe care s-au extras ştiri în respectiva săptămână, in functie de relevanta platformei pentru scopul activitatii – identificarea de semnale slabe, adica de stiri care anunta trenduri emergente.

Astfel, utilizatori puteau acorda următoarele scoruri: 1-platformă bună, 2-platformă medie(se poate păstra) sau 3 – platformă propusă spre eliminare.

Pentru fiecare platformă s-a calculat media, iar cele ce aveau scor de minim 2.5 urmau a fi mai atent analizate in fluxurile următoare, pentru a decide eliminarea sau pastrarea lor.

Raport voturi platforme pe liste de stiri

Lista de stiri: **Lista 14.08_21.08.2015;**

Generat de: **Radar Admin**

Sumar voturi:

Platforma	Numar voturi 1	Numar voturi 2	Numar voturi 3	Total voturi	Media
defensetech.org	0	1	0	1	2
defense-update.com	0	1	0	1	2
edition.cnn.com/TECH	0	4	2	6	2.33
http://abcnews.go.com/Technology	1	7	2	10	2.1
news.discovery.com	2	9	1	12	1.92
news.sciencemag.org	4	7	0	11	1.64
newsoffice.mit.edu	3	4	0	7	1.57
robohub.org	2	1	0	3	1.33
scitechdaily.com	4	6	0	10	1.6
searchenginewatch.com	0	5	3	8	2.38
smartcitiescouncil.com	0	5	2	7	2.29
spectrum.ieee.org	3	5	0	8	1.62
techcrunch.com	5	6	2	13	1.77
timesofindia.indiatimes.com/home/science	2	7	3	12	2.08
uncovercalifornia.com/business/technology	0	5	2	7	2.29
utsandiego.com/headlines/business/technology	1	2	1	4	2
www.3ders.org	2	7	1	10	1.9
www.3dprintingindustry.com	3	8	1	12	1.83
www.abc.net.au	0	3	4	7	2.57
www.aerospace-technology.com	0	7	1	8	2.12
www.aerotechnews.com	2	8	0	10	1.8
www.army-technology.com	0	3	2	5	2.4
www.asmarterplanet.com	0	2	0	2	2
www.augmented.org/blog	0	1	0	1	2
www.augmentedrealitytrends.com	0	2	3	5	2.6
www.azonano.com	4	4	1	9	1.67

Fig 6.1 Raport vot platforme

Raport voturi platforme pe liste de stiri

Lista de stiri: **Lista 09_16.10.2015;**

Generat de: **Radar Admin**

Sumar voturi:

Platforma	Numar voturi 1	Numar voturi 2	Numar voturi 3	Total voturi	Media
biopharma-asia.com	0	5	2	7	2.29
blogs.sap.com/innovation	0	6	3	9	2.33
campustechnology.com	0	6	3	9	2.33
defensetech.org	0	1	1	2	2.5
edition.cnn.com/TECH	0	6	1	7	2.14
http://abcnews.go.com/Technology	0	6	2	8	2.25
http://gadgets.ndtv.com/news	1	7	3	11	2.18
news.discovery.com	0	9	1	10	2.1
news.sciencemag.org	0	9	0	9	2
news.stanford.edu/news/socsci/	0	1	0	1	2
newsoffice.mit.edu	0	7	2	9	2.22
profit.ndtv.com/news/banking-finance	0	2	8	10	2.8
robohub.org	0	6	1	7	2.14
scitechdaily.com	1	6	0	7	1.86
searchenginewatch.com	0	3	2	5	2.4
smartcitiescouncil.com	0	2	0	2	2
spectrum.ieee.org	0	7	0	7	2
techcrunch.com	0	8	3	11	2.27
timesofindia.indiatimes.com/home/science	1	7	1	9	2
uncovercalifornia.com/business/technology	0	5	1	6	2.17
www.3ders.org	0	7	3	10	2.3
www.3dprintingindustry.com	0	7	2	9	2.22
www.abc.net.au	0	3	3	6	2.5

Raport voturi platforme pe liste de stiri

Lista de stiri: Lista 09_16.10.2015;

Generat de: Radar Admin

Sumar voturi:

Platforma	Numar voturi 1	Numar voturi 2	Numar voturi 3	Total voturi	Media
www.business-standard.com/technology-news	0	2	1	3	2.33
www.cbc.ca/news/technology	0	7	2	9	2.22
www.chinatechnews.com	0	1	2	3	2.67
www.cleantechnica.com	1	8	0	9	1.89
www.clickgreen.org.uk/rss/research	0	2	0	2	2
www.dailymail.co.uk/sciencetech	2	6	3	11	2.09
www.designboom.com/technology	1	0	0	1	1
www.economist.com/sections/science-technology	0	3	0	3	2
www.economist.com/topics/life-sciences	0	1	0	1	2
www.edsurge.com	0	5	2	7	2.29
www.electronicsweekly.com	0	5	0	5	2
www.environmentalleader.com	0	6	3	9	2.33
www.eschoolnews.com	0	3	1	4	2.25
www.euronews.com/sci-tech/	0	3	0	3	2
www.extremetech.com	0	9	0	9	2
www.firstpost.com/tech	0	6	4	10	2.4
www.foodnavigator-asia.com	0	4	3	7	2.43
www.foodproductiondaily.com	0	3	3	6	2.5
www.forumforthefuture.org	0	1	0	1	2
www.foxnews.com/tech	0	6	4	10	2.4
www.futurity.org	1	9	0	10	1.9
www.genengnews.com	1	8	1	10	2
www.gizmag.com	2	7	1	10	1.9
www.goodnewsnetwork.org	0	3	5	8	2.62
www.greentechmedia.com	0	6	1	7	2.14

Fig 6.2 Raport vot platforme

I.1.2. Repository de știri

Această aplicaţie/platformă conține toate ştirile existente în format excel şi txt. *Repository* este disponibilă la adresa: http://radarrepository.uefiscdi.ro/.

În perioada raportată au fost realizate următoarele funcționalități:

- Accesul pentru vizualizare pe bază de cont personalizat;
- Corelarea bazei de date Repository cu platforma TAGy în vederea uniformizării informațiilor.

Toate funcționalitățile acestei platforme în versiunea din perioada raportării și felul în care aceasta se corelează cu platforma TAGy sunt prezentate în ***Anexa 1_Repository_Tagy***.

I.2. Validarea umana (TAGy)

1. Consolidarea platformei electronice de identificare și clasificare a semnalelor slabe, în urma feedback-ului constant de la utilizatori/administratori
2. Procesul de evaluare și validare umană a știrilor

I.2.1. Consolidarea platformei electronice de identificare și clasificare a semnalelor slabe (TAGy)

In perioada de raportare s-a finalizat dezvoltarea, actualizarea si operationalizarea platformei electronice de evaluarea colaborativa a stirilor tehnologice**, Tagy** (**R7.1** tagy.uefiscdi.ro). Acesta a fost un proces continuu prin care echipa proiectului a integrat sugestiile venite din partea utilizatorilor si a rezolvat problemele tehnice aparute.

Astfel, in vederea eficientizarii procesului de validare umana, in cadrul platformei au fost dezvoltate cateva noi functionalitati (vizualizarea noilor aplicatii ale platformei este disponibila in ***Anexa 4_Platforma Tagy***):

- **Rapoarte**
  - **Raport general NS –** este la baza un raport general care ia in considerare doar non semnalele, si este folosit in vederea evaluarii activitatii de validare umana; ***Anexa 5_Raport general NS***
  - **Raport Vot Platforme –** Raport care scoate in evidenta o evaluare prin vot a celor mai bune platforme din punct de vedere al continutului stirilor. Procesul de votare s-a introdus tocmai pentru a primi un feedback din partea validatorilor cu privire la calitatea stirilor intrate in procesul de evaluare si implicit asupra surselor de provenienta. In functie de scorurile obtinute, echipa proiectului ia in calcul, impreuna cu alte criterii, eliminarea sau pastrarea sursei respective pentru fluxurile viitoare. **Anexa 6_Raport Vot Platforme**

- **Integrarea Tagy si Repository**
  - In perioada de raportare s-a finalizat interoperabilitatea dintre cele doua platforme; Astfel in platforma Repository se regasesc toate stirile colectate de-a lungul timpului prin diverse modalitati, inclusiv RSS feed (identificate unic printr-un cod txt, numele platformei de unde provine stirea, in format txt dar si xcel); Integrarea presupune ca alocarea stirilor in platforma Tagy se face automat din Repository – astfel ca exista o evidenta a stirilor care au intrat in fiecare saptamana in procesul de votare. ***Anexa 1_Repository_Tagy***

  - La sfarsitul proiectului, Tagy indeplineste toate conditiile si inglobeaza toate functionalitatile necesare pentru a fi o platforma operationala privind evaluarea colaborativa a stirilor tehnologice

I.2.2. Procesul de evaluare și validare umană a știrilor

In perioada iulie 2015 – noiembrie 2015, grupul de cunoastere tacita initial (CT7 din echipa proiectului) si-a pastrat acelasi rol de monitorizare si coordonare in procesul de validare umana pe platforma electronica Tagy.

Procesul de evaluare a stirilor a demarat cu alocarea catre evaluatorii umani a unui numar de 60 de texte in calitate de extractor si 60 de texte ca checker, pe care fiecare dintre cei 20 de colaboratori au trebuit sa le voteze intr-un interval de o saptamana. Lista de stiri alocate evaluatorilor insuma 1200 de texte selectionate din mai multe surse (platforme de stiri).

Ulterior, ritmul de lucru a crescut, listele alocate saptamanal contineau 2400 de texte(120 texte/extractor, 120 texte/checker pentru fiecare persoana), ajungandu-se la un maxim de 2500 texte evaluate saptamanal.

In urma unei evaluari intermediare realizate asupra procesului de validare umana, s-a luat decizia suplimentarii numarului de stiri alocate evaluatorilor; intr-o prima etapa s-a marit fluxul la 200 stiri/checker, 200 stiri/extractor iar ulterior s-a ajuns la 300/checker/extractor ceea ce a facut posibila evaluarea a cel putin 4000 stiri saptamanal la un numar de 18 evaluatori umani. Pana la sfarsitul perioadei de raportare au ramas 13 evaluatori, iar volumul de stiri evaluate a scazut proportional.

Pe toata perioada derularii proiectului au fost incarcate/ alocate/validate uman – 47 de liste de texte – adica un total de 149 717de stiri evaluate.

Pe tot parcursul acestui flux de lucru, echipa proiectului a monitorizat evolutia voturilor si a modului in care colaboratorii externi au reusit sa se integreze si sa participe la acest proces.

Monitorizarea s-a concretizat in analize si statistici relevante care scot in evidenta informatii relevante, atat din punct de vedere cantitativ (numarul de texte evaluate, numarul de semnale slabe sau nonsemnale votate, procentajul acestora in totalul stirilor, procentajul SS pe diferite surse, procentajul in functie de categoria aleasa a SS, etc) dar si calitativ (tipologia checker/extractor si modul cum acestia sunt fie autoritari/persuasivi, etc). Aceste analize au fost realizate pe baza seturilor de date de-a lungul desfasurarii activitatii in cadrul proiectului.

Pentru a scoate in evidenta cateva rezultate ale acestor statistici vom puncta faptul ca din cele **79997** de stiri evaluate in perioada de raportare, **2347** au fost votate ca semnale slabe, adica un procent de **3%** in medie. Acest procentaj a scazut in timp ca urmare a cresterii exigentelor cu care s-au tratat stirile, astfel ca daca in primele sesiuni procentul era undeva la 20%, apoi acesta a scazut la 16%, iar in prezent pe ultimele seturi de date ajungandu-se la 3%, si ca urmare la o medie generala sub 4%.Pentru informatii mai detaliate cu privire la statistici si analiza seturilor de date se poate accesa ***Anexa 7_Statistici si analize.actualizare.***

II. Implementarea componentei de analiză semantică din cadrul platformei de management de cunoaştere

II.1. Dezvoltarea scenariului de analiza semantică

Aceasta activitate a presupus crearea unor rețele semantice care să includă termenii specifici, cu alte cuvinte a unui dictionar cu structura arborescenta (**R7.5 Un scenariu de analiza semantica),** în programul Tropes Zoom. Aceste dezvoltari permit incadrarea automata a stirilor cu acuratete pe aceste taxonomii multiple, precum si vizualizarea lor multicriteriala.

Situația elaborării dicționarelor până la sfârșitul proiectului este detaliată în Anexa 8. Evidenta_domenii&tendinte.

În această perioadă au fost finalizate domeniile care au aveau dicţionare dezvoltate anterior: *Ageing, Electrical machinery, Environmental challenges, Environmental technology, Health, Medical equipment, Oceanography, Pharmaceutics, Shipbuilding, Space, Transport, Water industry.*

De asemenea au mai fost dezvoltate dicţionare pentru tendinţe mai nisate şi alte 7 domenii: 3d printer, additive manufacturing, artificial intelligence, artificial organ, big data, bioelectronics, biosensor, childcare, cloud computing, crowdsourcing, cybersecurity, driverless, energy storage, entertainment, fashion, fuel cell vehicles, gender, genomics, graphene,internet of things, migration, nuclear fusion, poverty, precision agriculture, quantum computing, recyclable thermoset plastic, renewable energy, retail, smart cities, sport, virtual reality, wearable technology, bionics, distributed manufacturing, precise genetic engineering.

Dicționarele pe domenii dar şi cel pentru tendinţe realizate în perioada iunie-noiembrie sunt integrate in versiunea Dicţionarelor Smart şi tendinţe din anexele: *Smart* din *Anexa 9 SmartDictionary_final* şi *Anexa 10. Dictionary tendinţe.*

II.2. Analiza automată

În perioada raportata, s-a finalizat dezvoltarea *in-house* a software-urilor necesare dezvoltarii componentei de analiza automata si semantica a platformei de management de cunoastere.

1. ***Software Dezvoltare Scenariu Semantic*** **(SDSS)**

A fost implementat sistemul informatic care permite clasificarea semantică a unui corpus mare de ştiri într-un număr de clustere la alegerea analistului uman. Pentru fiecare cluster, software-ul furnizează un set de cuvinte cheie, care vor fi incluse în dicționarul semantic. Acest software utilizează metoda de procesare a limbajului natural LDA (Latent Dirichlet Allocation).

Latent Dirichlet Analysis pentru clusterizarea nesupervizata a fluxului de stiri

Am implementat un proces generativ de clusterizare nesupervizata a fluxului de stiri pentru determinarea automata a similaritatilor computabile in corpus.

Intuitia principala a acestei tehnici este ca putem asocia in mod automat fiecarui cuvint o probabilitate de a semnala un tip de similaritate manifestata in corpus. Ceea ce se obtine este un vector de probabilitati pentru fiecare cuvint, dimensiunea vectorului fiind determinate de numarul de tipuri de similaritati considerate.

Inputul este consitutit de doua varabile:

Numarul de tipuri de similaritati dorite
Corpusul de analizat

In perioada de raportare am folosit 50 topice(aceasta repartizare a fost facuta pentru imbunatatirea predictiei de incadrare a stirilor) si *recurrent neural network* pentru definirea similaritatii la nivel de cuvinte. Aceasta clusterizare paralela a cuvintelor, independent de topice, este cuplata la inputul LDA pentru clusterizarea corpusului intr-o maniera care reduce fenomenul *data sparseness*. Rezultatul este o acoperire a corpusului mult mai mare, pentru aceeasi valoare a parametrului alfa.

Se maximeaza probabiliatea de ocurenta a unui cuvint tinind cont de contextul imediat inconjurator:
Corpusul nostru a fost format din peste 5000 de stiri text iar outputul a fost format din:

- Un set de cuvinte cheie pentru fiecare topic

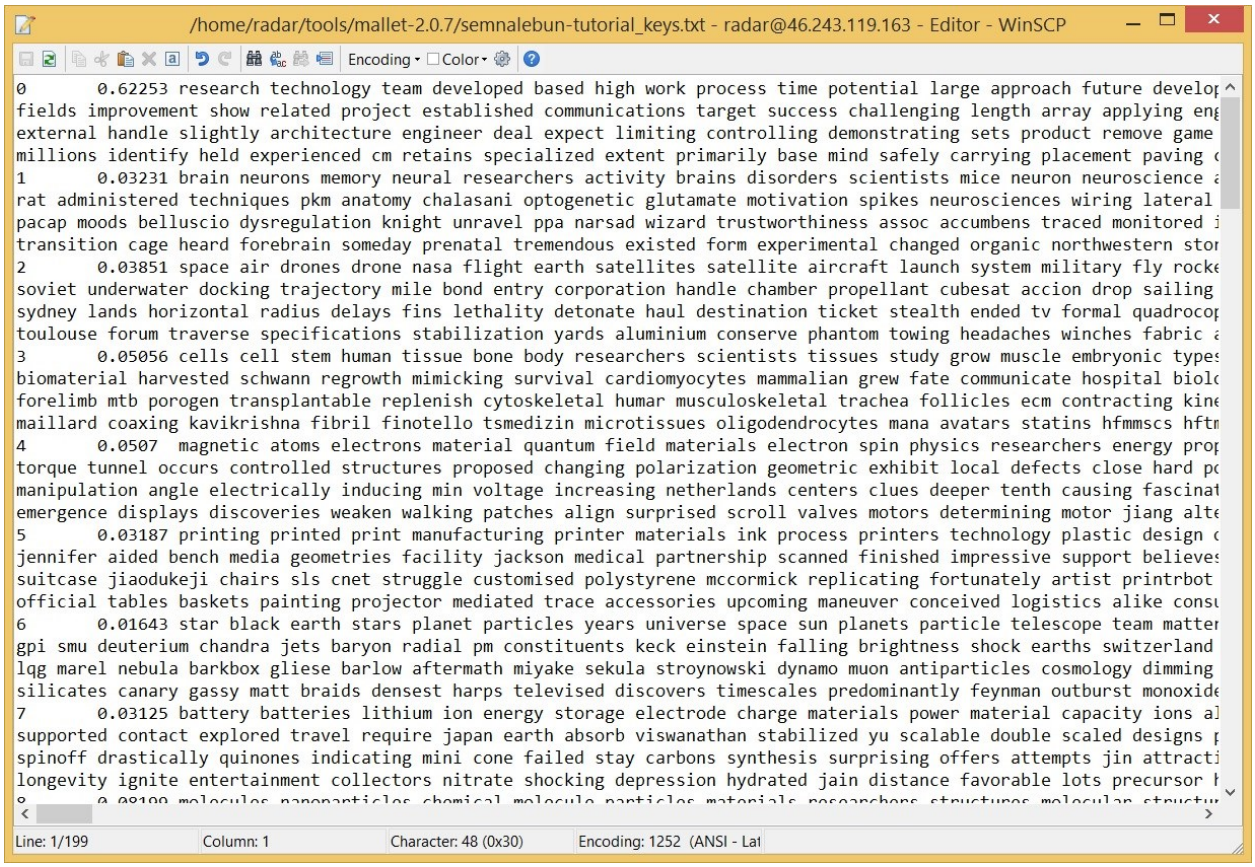


Fig. 7. Cuvintele cheie pentru fiecare cluser

Primul este numarul topicului (numerotarea incepe de la 0), iar al 2 numar reprezinta importanta acestuia.

- Repartizarea documentelor pe topice:

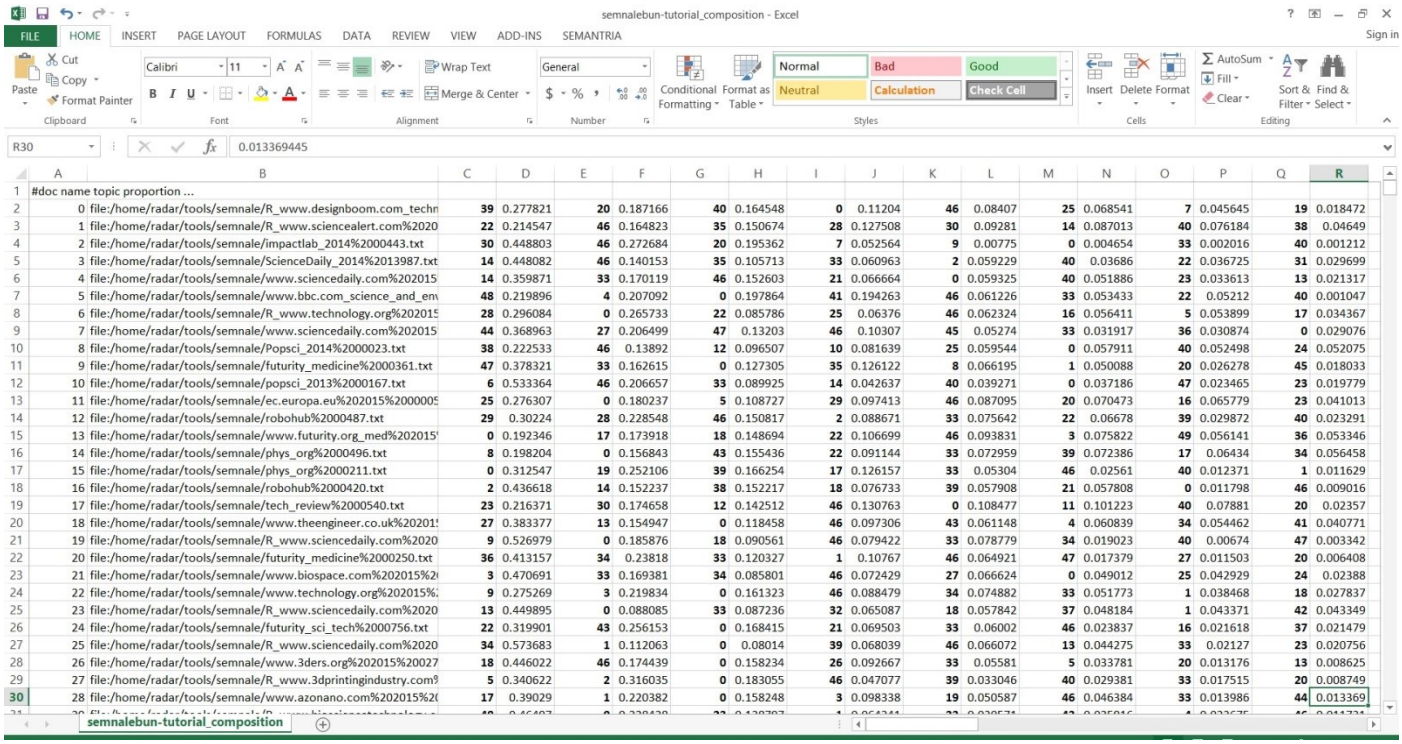


Fig. 8. Repartizarea documentelor pe topice

Fisierul 10 are topicul nr 6 ca topic principal cu 53%, topicul 46 cu 20%

2. *Software pentru Adnotarea Textelor* (SAT)

A fost dezvoltat un sistem de adnotare automată a ştirilor pe domenii descrise printr-un corpus de texte. Acest software utilizează tehnica SVM (Support Vector Machine) și permite ca, pornind de la un set de texte exemplu dintr-un domeniu (ex. foraj marin), să selecteze din corpusul de ştiri pe cele care se încadrează în acest domeniu (ştirile sunt ordonate în funcție de probabilitatea de încadrare). Dezvoltarea software-ului SAT are la baza “Data mining” – procesul de analiza a unor cantitati mari de date si de extragere a informatiilor relevante. Este un exercitiu interdisciplinar, in realizarea caruia, statistica, tehnologia bazelor de date, recunoasterea de tipare, inteligenta artificiala si vizualizarea isi au rolul lor

Pentru asta am folosit WEKA - o colectie de instrumente de vizualizare si algoritmi pentru analiza datelor si modelarea predictiva. Pentru a aplica acesti algoritmi pe corpusul nostru de stiri va trebui sa procesam datele.

Primul pas este sa transformam colectia de fisiere .txt intr-un format pe care weka il poate intelege. Textul va fi transformat in atribute cu ajutorul clasei StringToWordVector, folosind algoritmi de tokenizing si stemming.

Se va face o filtrare a atributelor folosind clasa Attribute Selection, pentru a le clasa in functie de evaluarile lor individuale (Ranker)

Urmatorul pas este alegerea unui clasificator, pentru a crea modelul de clasificare automata a stirilor. Support Vector Machine este o tehnica de clasificare bazata pe teoria invatarii statistice care a fost aplicata cu mult succes in multe probleme neliniare de clasificare si pentru multimi foarte mari de date. Idea algoritmului SVM este de a gasi un hiperplan care imparte optim setul de date de antrenament. Hiperplanul optim se poate distinge prin marginea de separare maxima dintre toate punctele de antrenare si hiperplan. Pentru o problema intr-un spatiu bidimensional algoritmul cauta dupa o dreapta care separa „cel mai bine" punctele din clasa pozitiva de punctele din clasa negativa. Hiperplanul este caracterizat printr-o functie de decizie de forma:

$$f(x) = \operatorname{sgn}\left(\langle \mathbf{w}, \mathbf{x} \rangle + b\right)$$

unde w este vectorul pondere, perpendicular pe hiperplan, "b" este un scalar care reprezinta marginea hiperplanului, "x" este esantionul curent testat si ·,· reprezinta produsul scalar. Sgn este functia semn care intoarce 1 daca valoarea obtinuta este mai mare sau egala cu 0 si -1 altfel. Daca w este de lungime 1, atunci <w, x> este de lungimea lui x de-a lungul directiei lui w. In general w va fi scalat prin ||w||.

In partea de antrenare algoritmul trebuie sa gaseasca vectorul normal "w" care conduce la cea mai mare margine "b" a hiperplanului. Problema pare foarte usor de rezolvat dar trebuie sa reamintim ca linia optima de clasificare trebuie sa clasifice corect toate elementele generate cu aceeasi distributie. Exista o multitudine de hiperplane care indeplinesc cerintele clasificarii, dar algoritmul incearca sa determine hiperplanul optim. Acest algoritm de invatare are loc intr-un spatiu in care exista produs scalar iar pentru datele liniar separabile construieste functia f din date empirice. Se bazeaza pe doua idei principale: dintre toate hiperplanele de separare exista un hiperplan optim unic care se distinge prin marginea maxima de separare dintre orice punct de antrenare si hiperplan. A doua idee este: capacitatea hiperplanului de a separa clasele descreste o data cu cresterea marginii. Pentru datele de antrenament care nu sunt liniar separabile printr-un hiperplan de separare in spatiul de intrare idea algoritmului SVM este de a proiecta datele de intrare intr-un spatiu de dimensiune mai mare prin intermediul unei functii Φ, si a incerca sa gaseasca acolo un hiperplan de separare cu marginea maxima. Aceasta conduce la o granita de separare neliniara in spatiul initial de intrare. Datorita faptului ca toti vectorii apar in produse scalare si prin utilizarea functiei nucleu ϕ(w),ϕ(x) ,este posibil sa calculam hiperplanul de separare fara a proiecta explicit datele in noul spatiu de trasaturi.

Pentru a gasi hiperplanul de separare optim care se distinge prin marginea maxima, trebuie sa rezolvam urmatoarea functie obiectiv:

$$\underset{\mathbf{w}\in H, b\in\Re}{\text{minimize}}\ \tau(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2$$

$$\textit{subject to}\ y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) \geq 1\ \textit{for all}\ i = 1,\ldots,m$$

Constrângerile asigura faptul ca $f(x_i)$ va fi +1 pentru y_i=+1 si -1 pentru y_i=-1. Aceasta problema este atractiva din punct de vedere al calculelor deoarece poate fi abordata prin rezolvarea unei probleme de programare patratica pentru care exista algoritmi eficienti. Functia τ se numeste functia obiectiv cu inegalitati de constrângeri. Acestea formeaza asa numita problema de optimizare primara. Pentru rezolvarea acestei probleme este mult mai convenabil sa lucram cu problema duala prin introducerea multiplicatorilor Lagrange $\alpha_i \geq 0$ si a Lagrangianului care conduce la problema duala de optimizare:

$$L(\mathbf{w},b,\alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{m}\alpha_i(y_i(\langle \mathbf{x_i}, w\rangle + b) - 1)$$

Cu multiplicatorii Lagrange $\alpha_i \geq 0$. Observam ca restrictiile sunt incluse in partea a doua a Lagrangianului si nu mai trebuie sa fie aplicate separat. Lagrangianul L trebuie sa fie maximizat in raport cu variabilele duale α_i, si minimizat in raport cu variabilele primare w si b. Aceasta conduce la:

$$\mathbf{w} = \sum_{i=0}^{m}\alpha_i y_i x_i$$

$$\sum_{i=0}^{m}\alpha_i y_i = 0$$

Vectorul solutie este o extensie in termeni de exemple de antrenament. Observam de altfel ca solutia w este unica (datorita convexitatii stricte pentru problema primara de optimizare), coeficienti α_i, nu trebuie sa fie unici. In acord cu teorema Karush-Kuhn-Tucker (KKT) doar multiplicatorii Lagrange α_i care sunt diferiti de zero sunt puncte de sprijin, corespunzatoare constrângerilor care se intâlnesc. Formal, pentru toti i=1,…,m acesta poate fi scris:

$$\alpha_i\left[y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1\right] = 0 \;\; for\;\; all\;\; i = 1,\ldots,m$$

Exemplele x_i pentru care $\alpha_i > 0$ se numesc vectori suport. Aceasta terminologie se refera la termeni corespondenti din teoria convexitatii. In acord cu conditiile KKT, acestia vor sta exact pe margine. Toate celelalte exemple de antrenament ramase sunt nerelevante. Prin eliminarea variabilelor primare w si b din Lagrangian obtinem asa numita problema de optimizare duala, care este problema care se rezolva de obicei in practica

$$\underset{\alpha\in\Re^m}{\text{maximize}}\, W(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m}\alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle$$

care se mai numeste si functia tinta, cu urmatoarele restrictii:

$$\text{subject to } \alpha_i \geq 0 \text{ for all } i = 1,\ldots, m \text{ and } \sum_{i=1}^{m} \alpha_i y_i = 0$$

Astfel hiperplanul poate fi scris, in problema de optimizare duala ca:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{m} y_i \alpha_i \langle x_i, x \rangle + b \right)$$

unde b este calculat utilizând conditiile KKT. Structura problemei de optimizare este foarte similara cu cea aparuta in formularea Lagrange mecanica. In rezolvarea problemei duala apare frecvent cazul in care doar o submultime de restrictii devine activa. De exemplu, daca vrem sa pastram o bila intr-o cutie atunci aceasta in mod uzual se va rostogoli intr-un colt. Restrictiile corespund peretilor care nu sunt atinsi de bila si care sunt nerelevanti in acel context deci pot fi eliminati. Totul a fost formulat in produse scalare. La nivel practic aceasta ofera posibilitatea ca algoritmul sa lucreze intr-un spatiu de dimensiune mare. Astfel noile sabloane $\Phi(x_i)$ pot fi rezultatul maparii datelor de intrare originale xi intr-un spatiu de dimensiune mai mare utilizând functia Φ. Maximizarea functiei tinta si evaluarea functiei de decizie implica calculul produsului scalar $\phi(x),\phi(x)$ intr-un spatiu de dimensiune mai mare. Aceste calcule costisitoare sunt reduse semnificativ prin utilizarea unui nucleu pozitiv definit k, astfel ca k(x, x') := x, x' . Aceasta substituire, care este referita uneori ca trucul nucleu, este utilizata pentru a extinde clasificarea cu hiperplane la SVM-ul neliniar. Trucul nucleu poate fi aplicat de vreme ce toti vectorii de trasaturi apar doar in produse scalare. Vectorii de trasaturi devin o expresie in spatiul de trasaturi si asadar Φ va fi functia prin care reprezentam vectorii de intrare in noul spatiu. Astfel functia de decizie va avea urmatoarea forma:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{m} y_i \alpha_i k(x, x_i) + b \right)$$

Principalul avantaj al acestui algoritm este ca nu necesita transpunerea tuturor datelor de intrare intr-un spatiu de dimensiune mai mare. De aceea nu vor exista nici calcule costisitoare ca in cazul retelelor neuronale. De asemenea obtinem o micsorare a setului de date in faza de antrenare când vom considera doar vectorii suport care de obicei sunt in numar redus. Un alt avantaj al acestui algoritm este ca permite utilizarea datelor de intrare cu un numar oricât de mare de trasaturi fara a creste exponential timpii de antrenare. Aceasta caracteristica nu este adevarata pentru retelele neurale, de exemplu algoritmul backprogapagion are probleme când trebuie sa lucreze cu un numnar mare de trasaturi. Singura problema care apare la SVM este numarul de vectori suport rezultati. Când numarul acestoar creste timpul de raspuns in faza de testare creste si el liniar.

Dupa alegerea corecta a parametrilor clasificatorului SVM, se va crea un model pornind de la un set de texte exemplu dintr-un domeniu. Pentru verificarea acuratetei modelului se va folosi cross-validation. Cross-validarea este o tehnica de validare pentru a evalua modul in care rezultatele unei analize statistice se va generaliza la un set de date independent . Acesta este utilizat in principal in cazul in care obiectivul este predictia , si se doreste estimarea cu exactitate modul in care un model predictiv va performa in practica.

Am folosit un corpus de training din 19 domenii. Dupa cross-validare s-a obtinut o acuratete de 93,70 % a modelului.

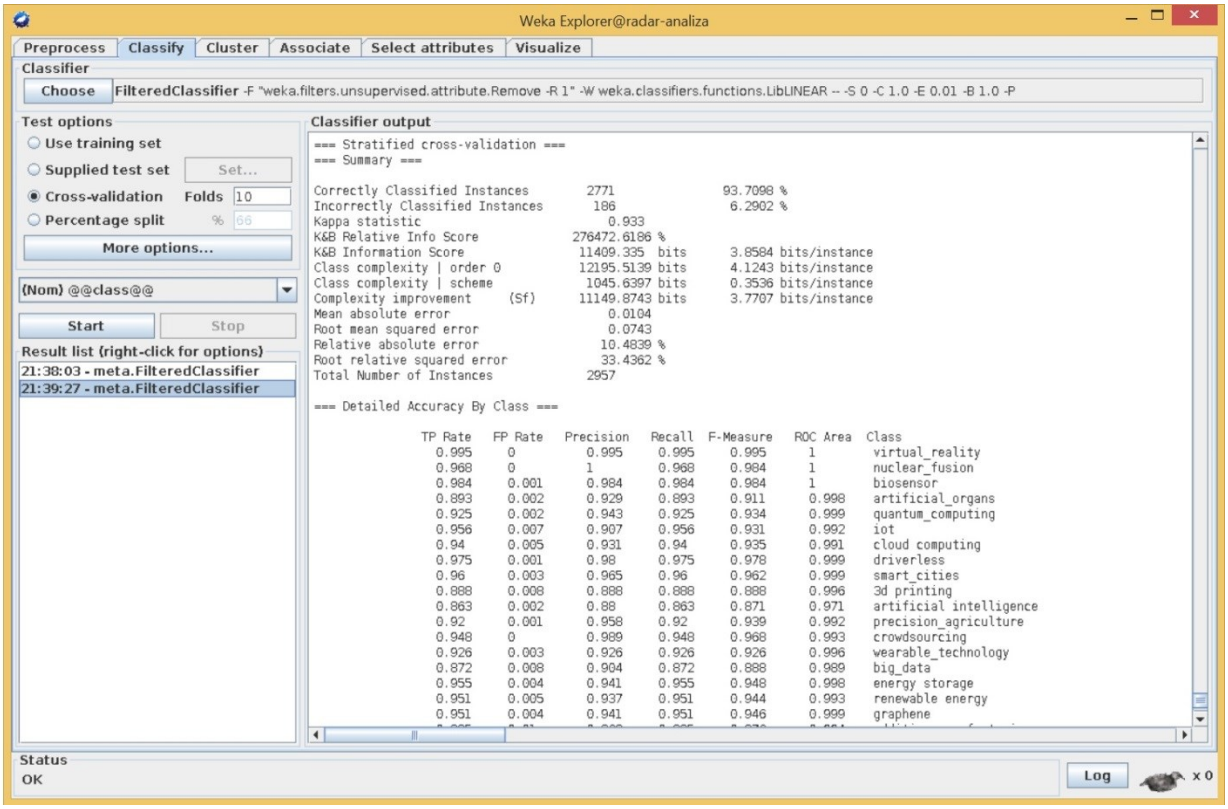


Fig. 9

Urmatorul pas este testarea unui corpus nou avand la baza modelul creat. Acesta este cel mai important pas deoarece scopul experimentului este ca modelul sa repartizeze corect stirile dintr-un corpus la prima vedere. Am folosit 400.000 de stiri iar outputul a fost :

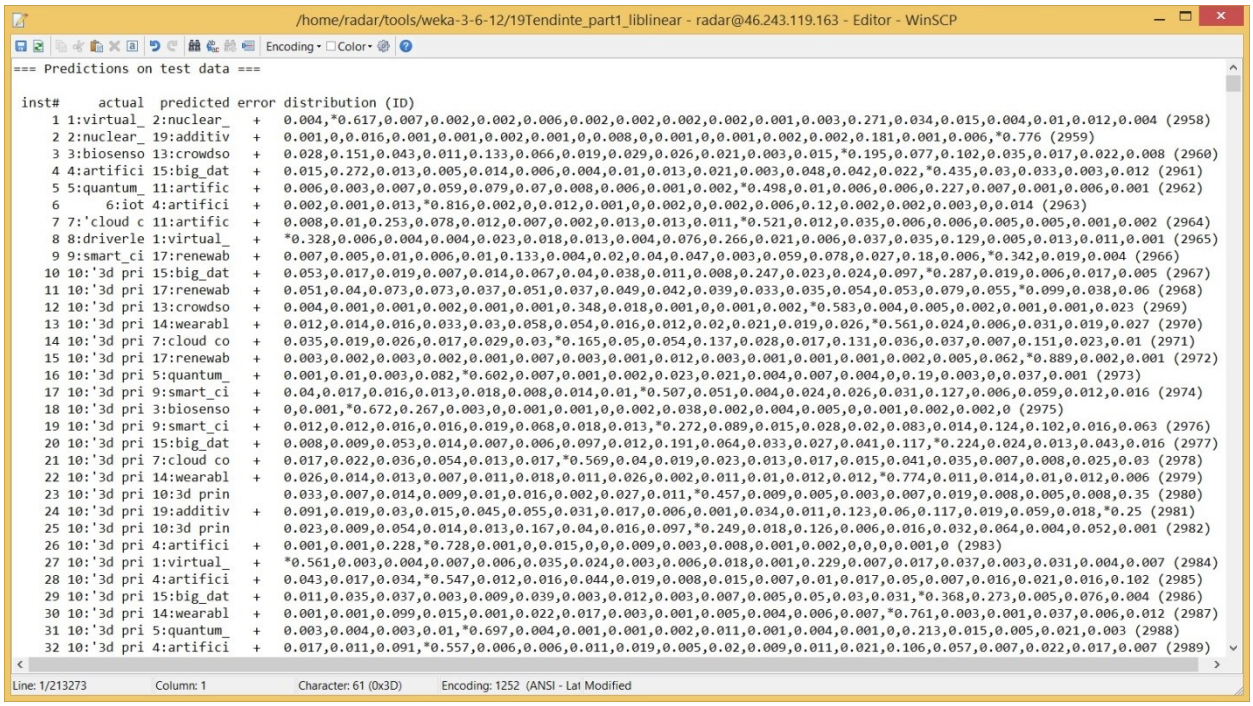


Fig. 10. Probabilitatea de incadrare a fiecarei stiri

Fiecare stire are un ID (numarul din paranteze de la finalul randului) pentru identificare. Dupa aplicarea unor functii logistice, L2 regularized logistic regression in cazul nostru, vom putea vedea probabilitatea de incadrare a fiecarei stiri.

Coloana a 2-a reprezinta domeniul corect al stirii, iar mai departe vedem probabilitatea de incadrare pentru fiecare domeniu.

Pentru o vizualizare mai buna, formatam aceste rezultate folosind un program perl si obtinem:

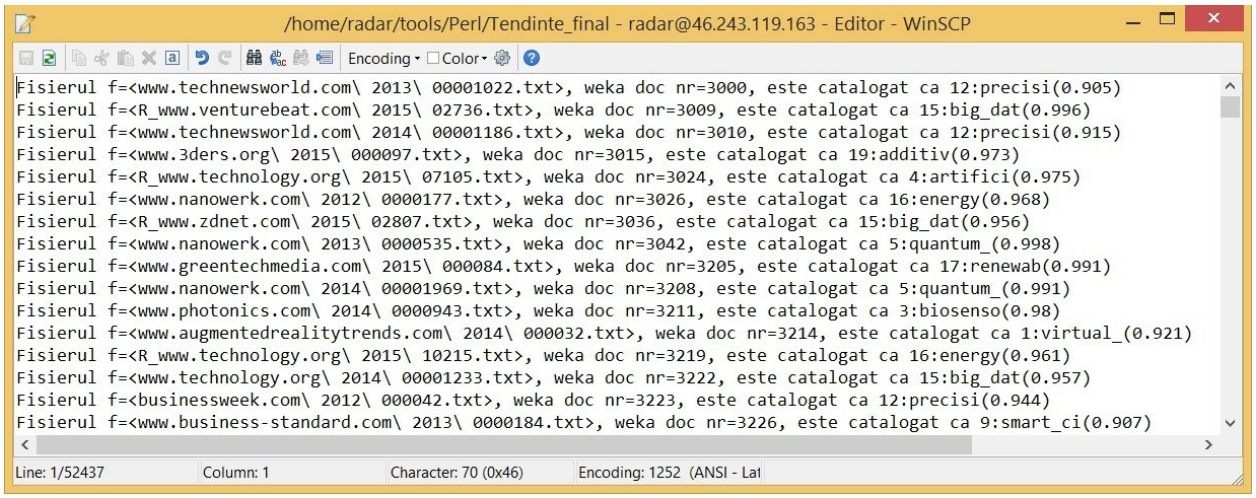


Fig. 11. Repartizarea stirilor pe domenii

Pentru determinarea pragului de acuratete folosim un procedeu de maximizare a tolerantei. Prin rezolvarea ecuatiilor de mai jos obtin un pran optimal pentru a asigura o calitate ridicata a clusterizarii:

$$P\left(\int_{-\infty}^{\bar{x}+ks} f(y)\,dy \geq 1-\beta\right)=1-\alpha \qquad P\left(\int_{-\infty}^{(\bar{x}+ks-\mu)/\sigma} \varphi(y)\,dy \geq 1-\beta\right)=1-\alpha$$

$$P\left((\bar{x}+ks-\mu)/\sigma \geq u_{1-\beta}\right)=1-\alpha$$

3. Folosirea Retelelor Neuronale pentru gasirea similaritatii

Clusterizarea automata a stiriilor, prezentata in sectiuna anterioara, este primul pas in gasirea similaritaiilor intre diferte parti constituent a unui flux de stiri. De la nivelul macro analiza, (document) se continua descendend catre nivelul de relational al cuvintelor. Focul este adus catre relatiile semantice dintre cuvinte.

In general o topica este descrisa de un set ce poate fi considerat din puct de vedere practic ca deschis. Cu alte cuvinte, un analizator trebuie sa aiba capaciatea sa recunoasca o topica chiar daca cuvinte noi sint folosite. Acest lucru se realizeaza prin determinarea similaritii semantice intre cuvinte. In acest scop se folosesc

modele neurale care sint capabile sa atinga un grad ridicat de acuratete in prezicerea similaritatii individuale (ca opusa similaritatii combinatoriale).

Determinarea similaritatii se face prin calcul vectoriala. Contextul este representat ca vectori iar similaritatea este calculate via proximitate vectoriala data de o metrica oarecare. Specific, in acest caz, folosim produsul punctual si metrica euclidiana.

Folosim procul “sac-de-cuvinte”. Peste corpusul de stiri, calculam si maximizam urmatoarea serie de probabilitati:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}\log p(w_{t+j}|w_t)$$

Deci vom gasi setul de features care maximizeaza probabilitatea de mai sus.

Acest process este unul iterative. Optimizarea este obinuta pas cu pas, repetind calcularea max loglikehood de mai sus.

In figura urmatoare se prezinta arhitectura retelei neurale recurente.

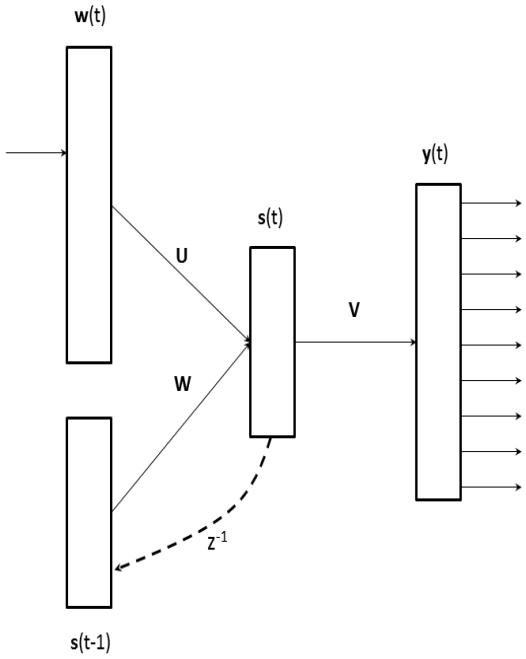


Fig. 12.

Formula de liniarizarea a inputului este:

$$\mathbf{s}(t) = f\left(\mathbf{U}\mathbf{w}(t) + \mathbf{W}\mathbf{s}(t-1)\right) \qquad (1)$$

$$\mathbf{y}(t) = g\left(\mathbf{V}\mathbf{s}(t)\right), \qquad (2)$$

where

$$f(z) = \frac{1}{1+e^{-z}}, \quad g(z_m) = \frac{e^{z_m}}{\sum_k e^{z_k}}. \qquad (3)$$

In acest fel obtinem un set de cuvinte similare pentru fiecare target de care avem nevoie. Prezentam citeva exemple mai jos.

Proiect cofinanţat din Fondul Social European, prin Programul Operaţional "Dezvoltarea Capacităţii Administrative", în perioada 2007-2013

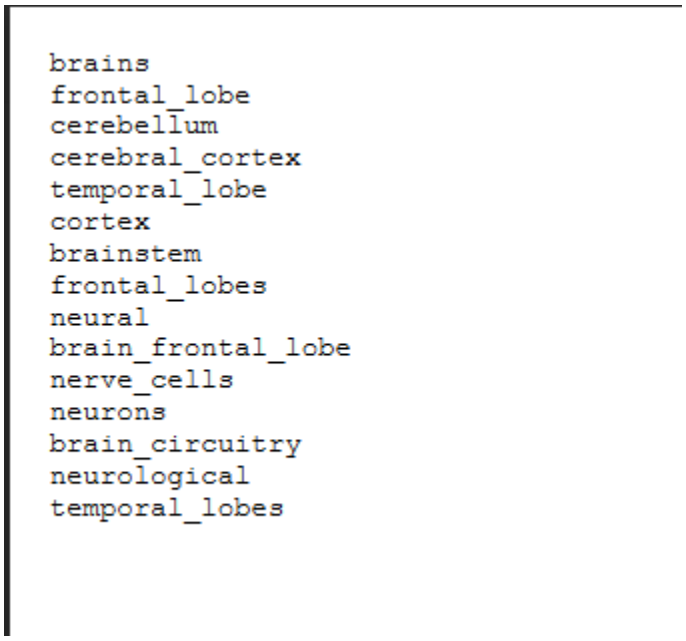


Fig. 13. Multimea de Similaritate pentru brain

Un system care detecteaza semnale slabe vs. non-semnale pentrurapoarte asupra activitatiilor de cercetare ale creierului, va fi activat nunumai de cuvintul “barain”, dar si de cuvinte ca “nerve_cells”, sau “neurons”.

Probabilitatiile apriori sint calculate de RNR (reteua neuronala recurenta). Aceste probabilitati sint apriori pentru ca au fost calculate pentru cuvinte , intr-un mod individual. Ele reprezinta doar asocieri unul la unul intre cuvinte. Semantica unei topici se calculeaza via nivelul de discriminativitate intre topici , peste nivelul individual.

De exemplu , pentru similaritatiile individuale obtinute in auput de RNR pentru “dentist”, vezi figura …, avem

```
orthodontist
Dentist
periodontist
doctor
pediatric_dentist
dentists
oral_surgeon
dental
dermatologist
plastic_surgeon
endodontist
cosmetic_dentist
chiropractor
dental_hygienist
dentistry
```

Fig. 14. Similaritati individuale dentist

"plastic surgeron" , poate intr-adevar activa in mod corect the modulu; care cauta semnale slabe pentru "dentist" . Dar, aceasta probabilitate trebuie ponderata via probabiltatii globale.

Cu o technica de redistribuire a probabilitatiilor, rezolvind si problema probabilitatiilor asociarilor cu cuvinte inca neprocesate, folosind o formula, (Katz):

$$P_{bo}(w_i \mid w_{i-n+1} \cdots w_{i-1})$$

$$= \begin{cases} d_{w_{i-n+1} \cdots w_i} \dfrac{C(w_{i-n+1} \cdots w_{i-1} w_i)}{C(w_{i-n+1} \cdots w_{i-1})} & C(w_{i-n+1} \cdots w_i) > k \\ \alpha_{w_{i-n+1} \cdots w_{i-1}} P_{bo}(w_i \mid w_{i-n+2} \cdots w_{i-1}) \end{cases}$$

Cu C reprezentind numarul de ocurente pentru o anumita topica.

4. Annotarea Elementelor Semnificative

In detectarea semnalelor slabe vs. tari, nu numai cuvintele cheie si cele similare lor sunt importante, sau domeniul si conectivitatiile semantice gasite prin procedeele prezentate in sectiunile de mai sus.

In aceasta sectiune vom prezenta modulul de annotare a elementelor semnificative dintr-un document. Dupa identificarea cuvintelor cheie si incadrarea intr-un topic a unui document, acest modul detecteaza agentii si relatiile dintre cuvintele de interes si agentii respective.

La baza acestui modul sta un model al limbajului generat prin tehnica Named Entity Recognition (NER), sau Identificarea Entitatiilor Semnificative (IES). Tehnica IES a fost dezvoltata pe baza algoritmilor HMM and SVM.

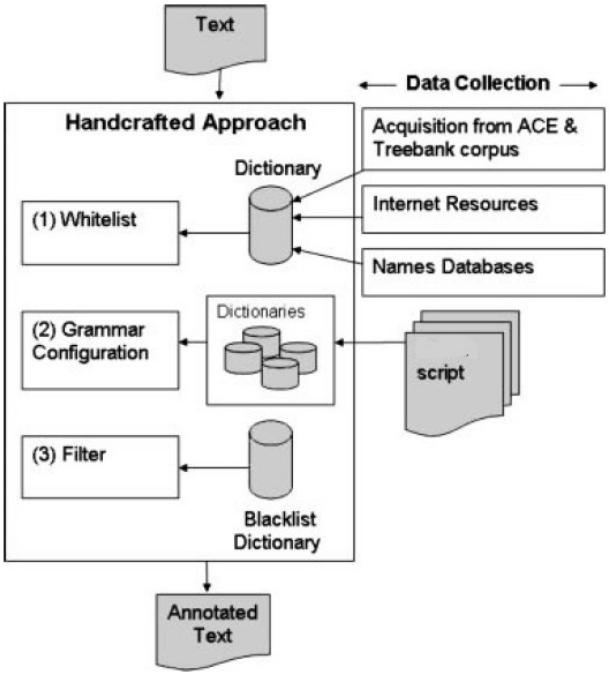


Fig. 15.

Modelul general este prezentat in *Figura de mai sus* Detectarea agentilor si tipului este un process in care se creaza un model, care este folosit pentru annotarea agentilor in documentele de interes. Formula este data via un proces Gibbs (lanturi Markov)

$$\mathrm{P}_A(\mathbf{s}^{(t)}|\mathbf{s}^{(t-1)}) = \frac{\mathrm{P}_M(s_i^{(t)}|\mathbf{s}_{-i}^{(t-1)}, \mathbf{o})^{1/c_t}}{\sum_j \mathrm{P}_M(s_j^{(t)}|\mathbf{s}_{-j}^{(t-1)}, \mathbf{o})^{1/c_t}}$$

Ceea ce se obtine in output este:

Researchers at UC San Diego School of Medicine conducted the first population-based study that characterizes the associatio n and temporal relationship between gastrointestinal stromal tumors (GIST) and other cancers. The results indicate that one in 5.8 patients with GIST will develop additional malignancies before and after their diagnosis.Specifically, patients wit h GIST are more likely to develop other sarcomas, non-Hodgkin's lymphoma, carcinoid tumors, melanoma, colorectal, esophageal, pancreatic, hepatobiliary, non-small cell lung, prostate and renal cell cancers."Only 5 percent of patients with gastroi ntestinal stromal tumors have a hereditary disorder that predisposes them to develop multiple benign and malignant tumors," said Jason K. Sicklick, MD, assistant professor of surgery and UC San Diego Moores Cancer Center surgical oncologist. "The research indicates that these patients may develop cancers outside of these syndromes, but the exact mechanisms are not ye t known."The researchers said further studies are needed to understand the connection between GIST and other cancers, but t he findings may have clinical implications."Patients diagnosed with gastrointestinal stromal tumors may warrant considerati on for additional screenings based on the other cancers that they are most susceptible to contract," said co-author James D . Murphy, MD, assistant professor of radiation oncology and UC San Diego Moores Cancer Center radiation oncologist.When com pared to the United States population, the researchers found that people with GIST had a 44 percent increased prevalence of cancers occurring before a GIST diagnosis and a 66 percent higher risk of developing cancers after diagnosis. The most com mon tumors were those of the genitourinary tract, breast, respiratory and blood.Non-Hispanic patients had a higher incidenc e of other cancers before a GIST diagnosis. Patients whose tumors were smaller than 10 centimeters had a higher probability of a second cancer than patients whose growth was larger. People with tumors smaller than 2 cm had the greatest likelihood of developing additional malignancies, both before and after diagnosis.

Fig. 16

Putem astfel sa punem in relatie nu numai substantivele commune dar si substantivele proprii care definesc agentii care sint mentionati in document.

In exemplul de mai sus reusim sa extragem automat urmatoarele fapte:

GIST este un tip de tumoare (gastrointestinal)

San Diego este un centru de cercetare responsabil de o descoperire cu implicatii importante

D Murph este un cercetator Implicat in GIST

In viitor, adica in prelucrarea documentelor ulterioare, putem urmarii ceea ce se intampla cu agentii, stiind ca ei sint conectacti la o anumita topica. Astfel se asigura un control al acuratetei procesarii semnalelor , mult mai ridicat.

Concluzii

Am dezvoltat un sistem de procesare a semnalelor pe baza a patru instrumente fundamentale:

- SVM
- LDA
- RNR
- NER

Rezulatele au confirmat puterea de procesare ridicata a sistemului implemetat.

Prezentarea algoritmilor utilizati pentru clasificarea si adnotarea automata a stirilor:

I. LDA:

1. Se instaleaza softul MALLET - http://mallet.cs.umass.edu/
2. Urmatorul pas este importarea datelor – acestea se pun intr-un singur folder.

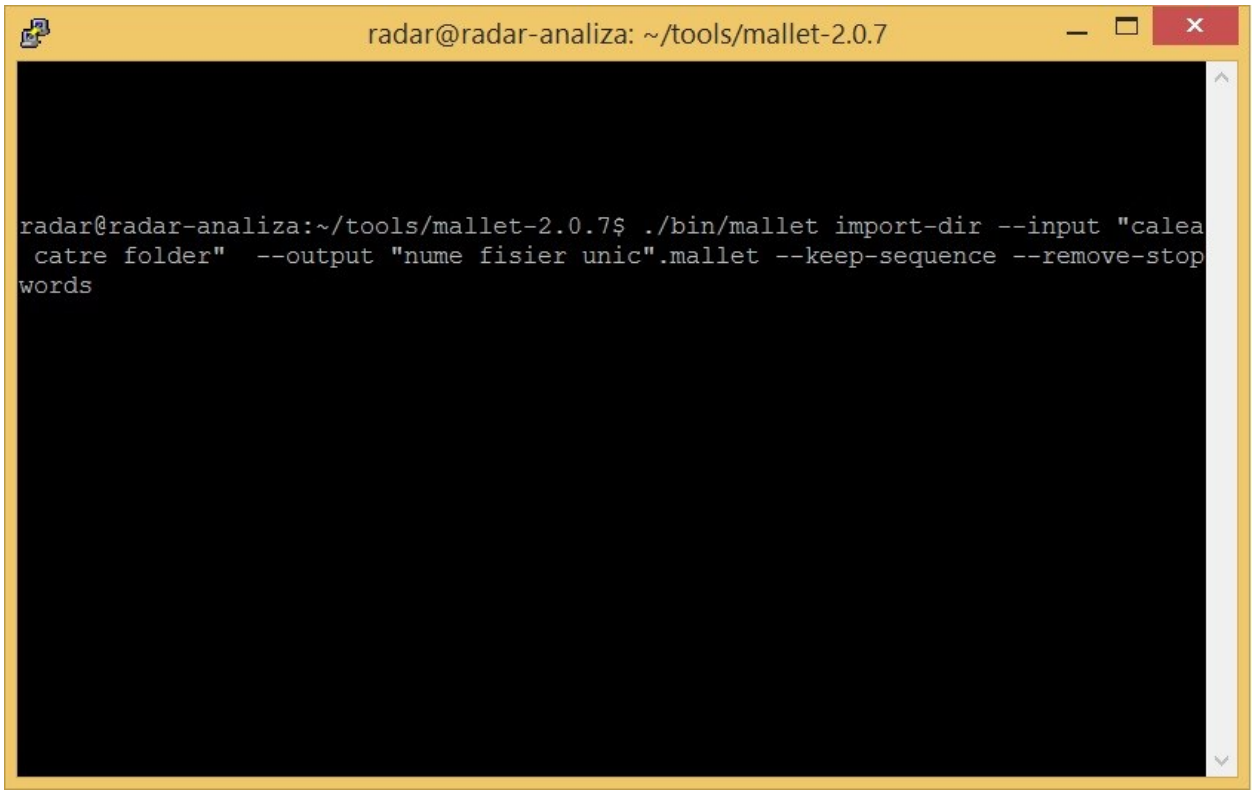


Fig. 16

Pentru a lucra cu intregul corpus va trebui sa transformam toate fisierele intr-un singur fisier de tip .mallet. Pentru acest pas folosim comanda **import**

Comanda –keep-sequence pastreaza ordinea intiala a fisierelor, iar –remove-stopwords sterge cuvintele care ar obstructiona analiza corecta.

3. Apoi se construieste modelul folosind comanda **train topic**

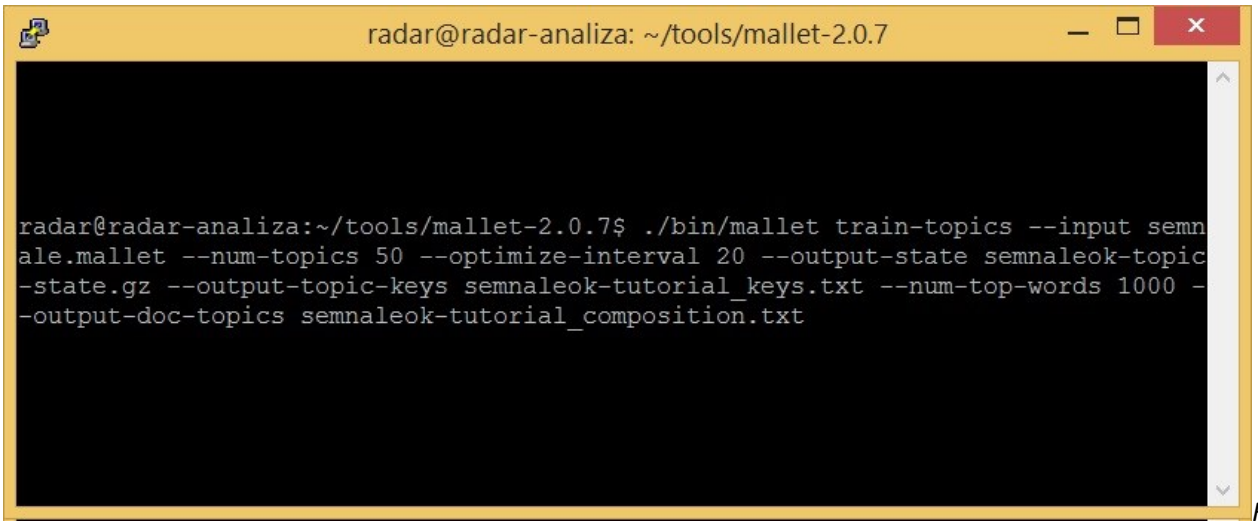


Fig. 17

4. Aceasta comanda:
    - Deschide fisierul *.**mallet**
    - Mallet creeaza 50 de topice
    - Creeaza un fisier txt ce contine toate cuvintele cheie pentru fiecare topic

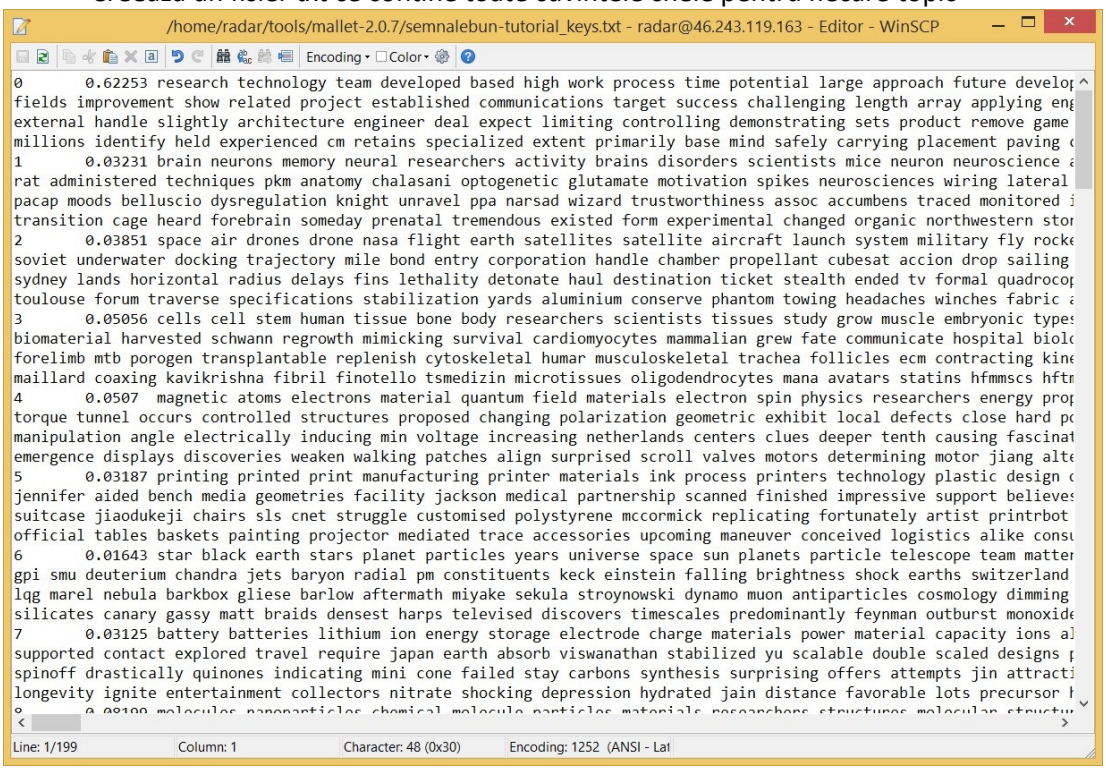
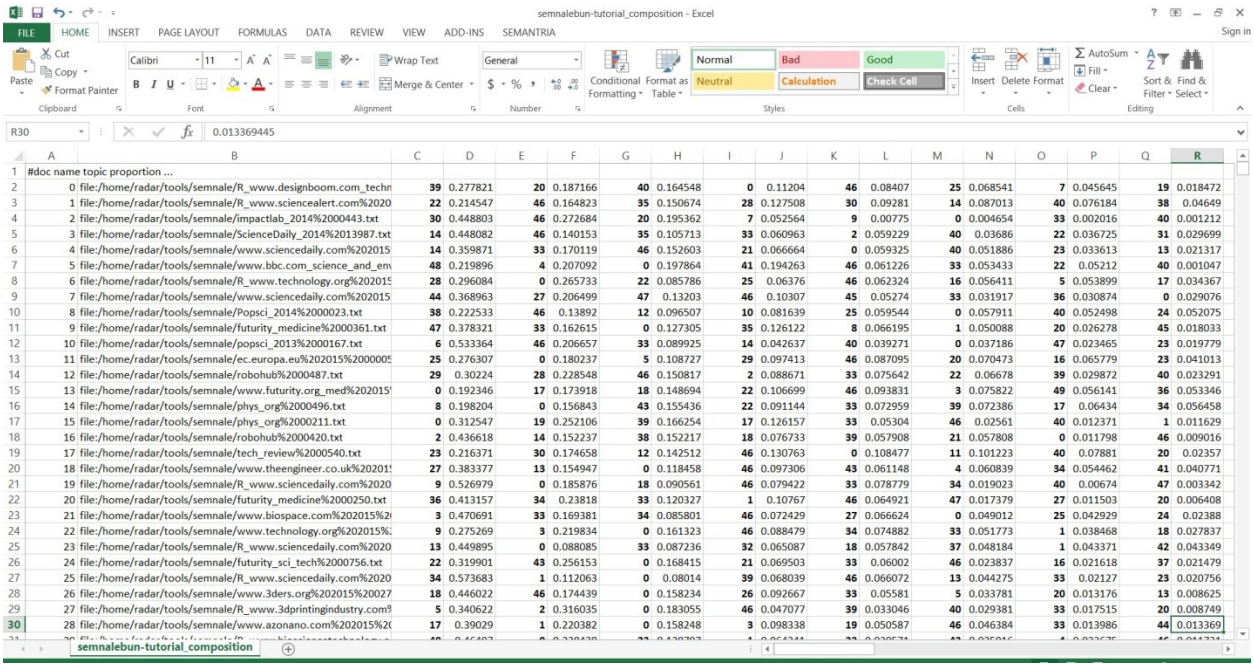


Fig. 18

5. Deschidem fisierul *-**tutorial_composition.txt** intr-un fisier nou Excel pentru a-l putea vizualiza corect. Acesta contine repartizarea fiecarui fisier pe topice



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	#doc	name	topic	proportion	...													
2	0	file:/home/radar/tools/semnale/R_www.designboom.com_techn	39	0.277821	20	0.187166	40	0.164548	0	0.11204	46	0.08407	25	0.068541	7	0.045645	19	0.018472
3	1	file:/home/radar/tools/semnale/R_www.sciencealert.com%2020	22	0.214547	46	0.164823	35	0.150674	28	0.127508	30	0.09281	14	0.087013	40	0.076184	38	0.04649
4	2	file:/home/radar/tools/semnale/impactlab_2014%2000443.txt	30	0.448803	46	0.272684	20	0.195362	7	0.052564	9	0.00775	0	0.004654	33	0.002016	40	0.001212
5	3	file:/home/radar/tools/semnale/ScienceDaily_2014%2013987.txt	14	0.448082	46	0.140153	35	0.105713	33	0.060963	2	0.059229	40	0.03686	22	0.036725	31	0.029699
6	4	file:/home/radar/tools/semnale/www.sciencedaily.com%202015	14	0.359871	33	0.170119	46	0.152603	21	0.066664	0	0.059325	40	0.051886	23	0.033613	13	0.021317
7	5	file:/home/radar/tools/semnale/www.bbc.com_science_and_en	48	0.219896	4	0.207092	0	0.197864	41	0.194263	46	0.061226	33	0.053433	22	0.05212	40	0.001047
8	6	file:/home/radar/tools/semnale/R_www.technology.org%202015	28	0.296084	0	0.265733	22	0.085786	25	0.06376	46	0.062324	16	0.056411	5	0.053899	17	0.034367
9	7	file:/home/radar/tools/semnale/www.sciencedaily.com%202015	44	0.368963	27	0.206499	47	0.13203	46	0.10307	45	0.05274	33	0.031917	36	0.030874	0	0.029076
10	8	file:/home/radar/tools/semnale/Popsci_2014%2000023.txt	38	0.222533	46	0.13892	12	0.096507	10	0.081639	25	0.059544	0	0.057911	40	0.052498	24	0.052075
11	9	file:/home/radar/tools/semnale/futurity_medicine%2000361.txt	47	0.378321	33	0.162615	0	0.127305	35	0.126122	8	0.066195	1	0.050088	20	0.026278	45	0.018033
12	10	file:/home/radar/tools/semnale/popsci_2013%2000167.txt	6	0.533364	46	0.206657	33	0.089925	14	0.042637	40	0.039271	0	0.037186	47	0.023465	23	0.019779
13	11	file:/home/radar/tools/semnale/ec.europa.eu%202015%2000005	25	0.276307	0	0.180237	5	0.108727	29	0.097413	46	0.087095	20	0.070473	16	0.065779	23	0.041013
14	12	file:/home/radar/tools/semnale/robohub%2000487.txt	29	0.30224	28	0.228548	46	0.150817	2	0.088671	33	0.075642	22	0.06678	39	0.029872	40	0.023291
15	13	file:/home/radar/tools/semnale/www.futurity.org_med%202015	0	0.192346	17	0.173918	18	0.148694	22	0.106699	46	0.093831	3	0.075822	49	0.056141	36	0.053346
16	14	file:/home/radar/tools/semnale/phys_org%2000496.txt	8	0.198204	0	0.156843	43	0.155436	22	0.091144	33	0.072959	39	0.072386	17	0.06434	34	0.056458
17	15	file:/home/radar/tools/semnale/phys_org%2000211.txt	0	0.312547	19	0.252106	39	0.166254	17	0.126157	33	0.05304	46	0.02561	40	0.012371	1	0.011629
18	16	file:/home/radar/tools/semnale/robohub%2000420.txt	2	0.436618	14	0.152237	38	0.152217	18	0.076733	39	0.057908	21	0.057808	0	0.011798	46	0.009016
19	17	file:/home/radar/tools/semnale/tech_review%2000540.txt	23	0.216371	30	0.174658	12	0.142512	46	0.130763	0	0.108477	11	0.101223	40	0.07881	20	0.02357
20	18	file:/home/radar/tools/semnale/www.theengineer.co.uk%20201	27	0.383377	13	0.154947	0	0.118458	46	0.097306	43	0.061148	4	0.060839	34	0.054462	41	0.040771
21	19	file:/home/radar/tools/semnale/R_www.sciencedaily.com%2020	9	0.526979	0	0.185876	18	0.090561	46	0.079422	33	0.078779	34	0.019023	40	0.00674	47	0.003342
22	20	file:/home/radar/tools/semnale/futurity_medicine%2000250.txt	36	0.413157	34	0.23818	33	0.120327	1	0.10767	46	0.064921	47	0.017379	27	0.011503	20	0.006408
23	21	file:/home/radar/tools/semnale/www.biospace.com%202015%2	3	0.470691	33	0.169381	34	0.085801	46	0.072429	27	0.066624	0	0.049012	25	0.042929	24	0.02388
24	22	file:/home/radar/tools/semnale/www.technology.org%202015%	9	0.275269	3	0.219834	0	0.161323	46	0.088479	34	0.074882	33	0.051773	1	0.038468	18	0.027837
25	23	file:/home/radar/tools/semnale/R_www.sciencedaily.com%2020	13	0.449895	0	0.088085	33	0.087236	32	0.065087	18	0.057842	37	0.048184	1	0.043371	42	0.043349
26	24	file:/home/radar/tools/semnale/futurity_sci_tech%2000756.txt	22	0.319901	43	0.256153	0	0.168415	21	0.069503	33	0.06002	46	0.023837	16	0.021618	37	0.021479
27	25	file:/home/radar/tools/semnale/R_www.sciencedaily.com%2020	34	0.573683	1	0.112063	0	0.08014	39	0.068039	46	0.066072	13	0.044275	33	0.02127	23	0.020756
28	26	file:/home/radar/tools/semnale/www.3ders.org%202015%20027	18	0.446022	46	0.174439	0	0.158234	26	0.092667	33	0.05581	5	0.033781	20	0.013176	13	0.008625
29	27	file:/home/radar/tools/semnale/R_www.3dprintingindustry.com%	5	0.340622	2	0.316035	0	0.183055	46	0.047077	39	0.033046	40	0.029381	33	0.017515	20	0.008749
30	28	file:/home/radar/tools/semnale/www.azonano.com%202015%2	17	0.39029	1	0.220382	0	0.158248	3	0.098338	19	0.050587	46	0.046384	33	0.013986	44	0.013369

Fig. 19

Exemplu: Fisierul 10 are topicul nr 6 ca topic principal cu 53%, topicul 46 cu 20%.

II. SVM:

1. Se instaleaza WEKA, un software “open-source” de invatare automata si LibSVM, un calsificator SVM
2. Cele 2 corpusuri de texte, cel de invatare si cel de test vor fi transformate in fisiere ARFF (**Attribute-Relation File Format**). Fiecare text va fi reprezentat ca un rând in noul fisier. Pentru asta vom folosi un program Java, numit TextDirectoriestoARFF. Pentru a utiliza acest program, fiecare categorie din colecția de texte trebuie să aibă propriul său director.

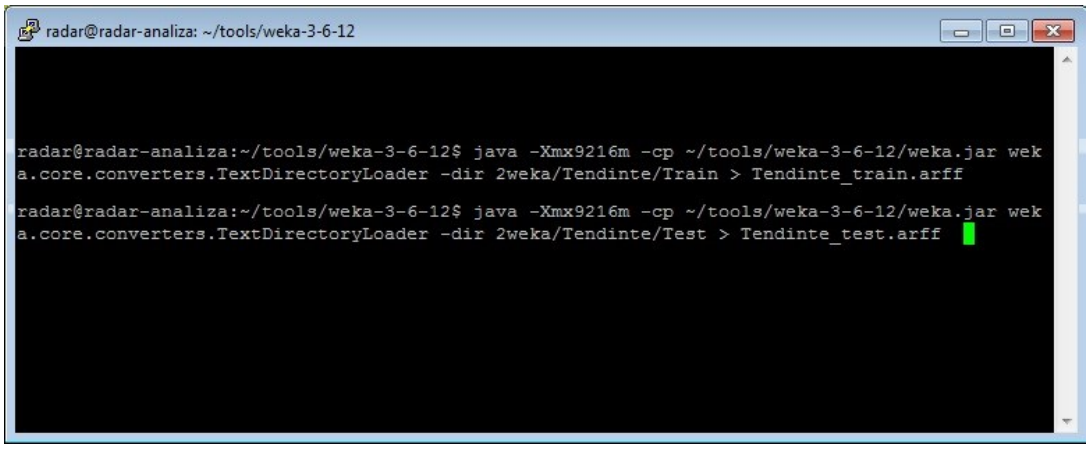


Fig. 21

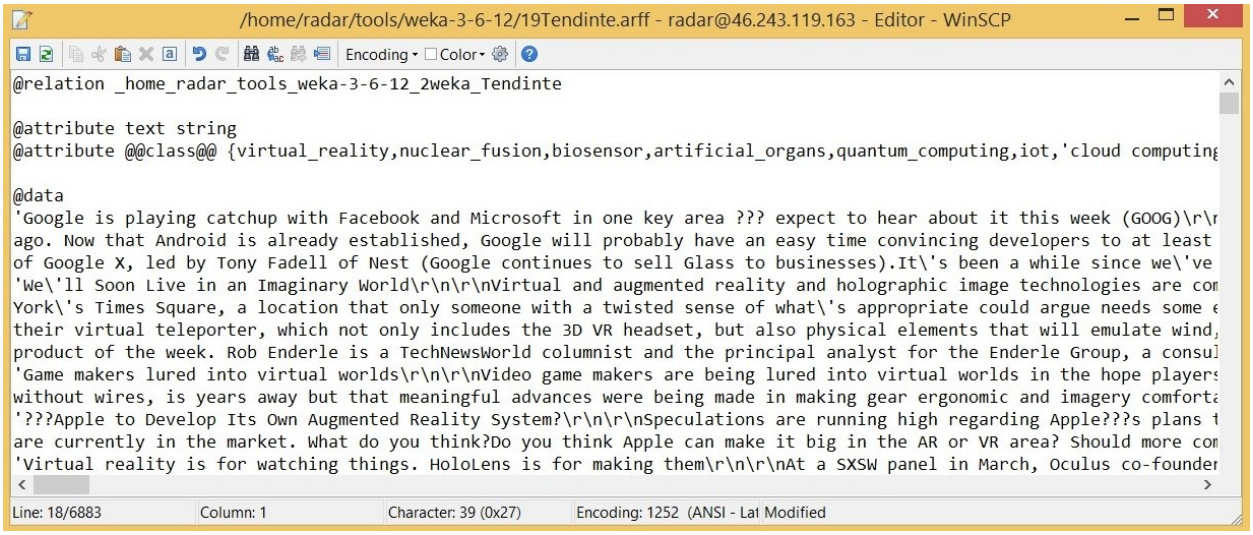


Fig. 22

Iar outputul va arata astfel

3. Convertirea atributelor de tip “string” in atribute numerice

   Fisierul de la pasul anterior are doar 2 atribute: “class” – care reprezinta categoriile corpusului si “text” – continutul textelor. Trebuie să transformam acest fișier într-un format de unde putem extrage caracteristici (atribute) și sa avem valori numerice pentru ele.

   Pentru asta vom folosi tot un program java numit **StringtoWordVector**

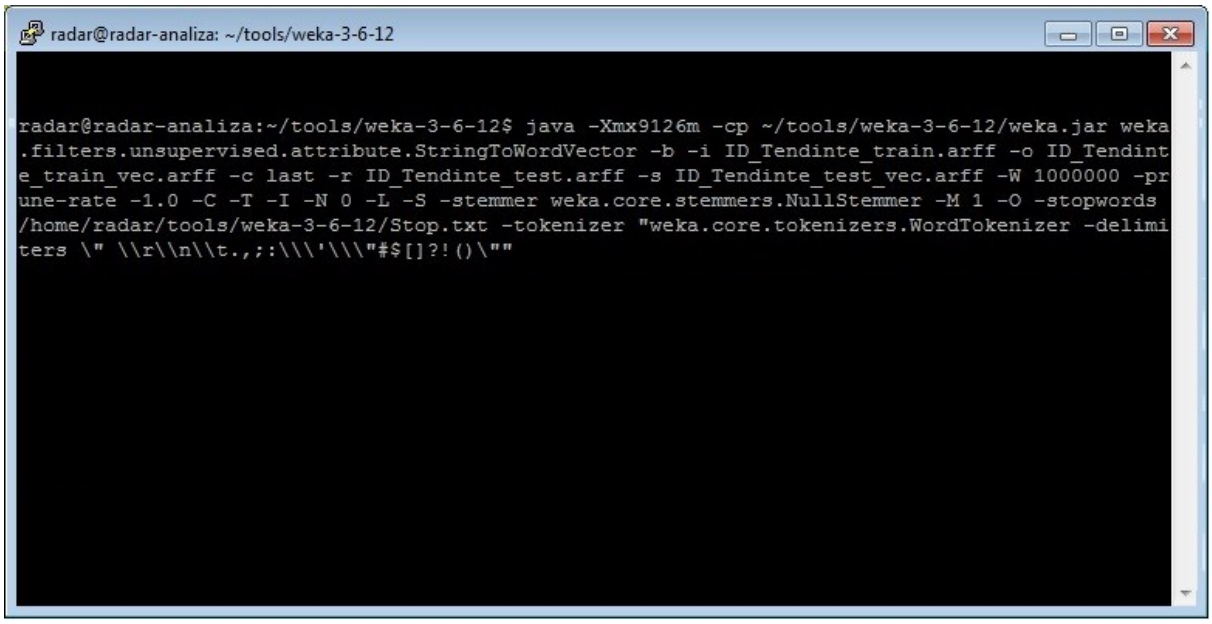


Fig. 23

Rezultand fisierul de mai jos

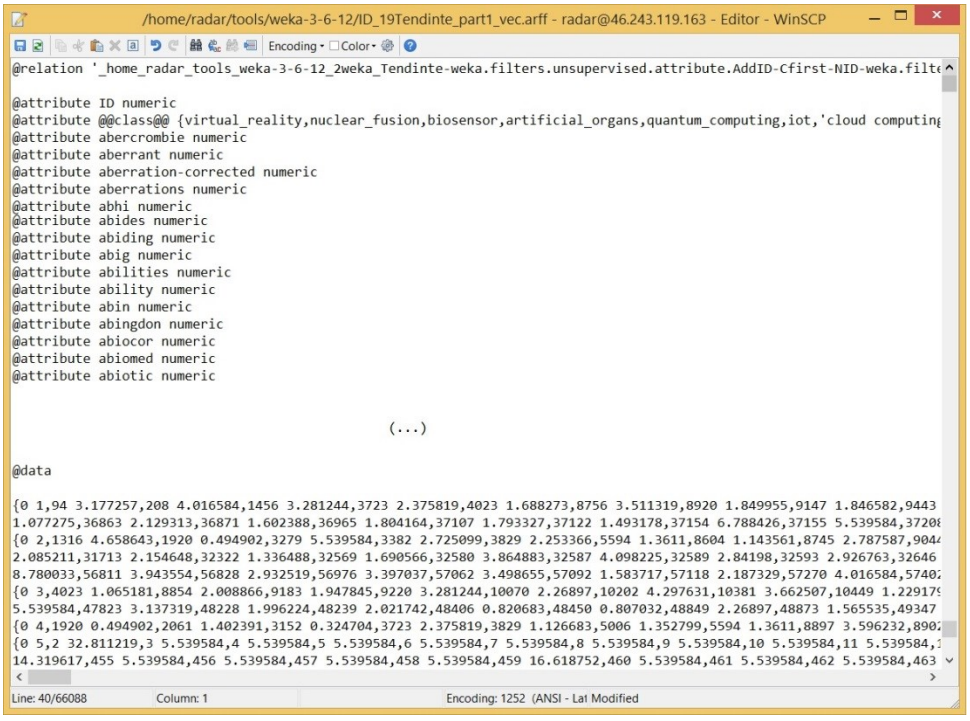


Fig. 24

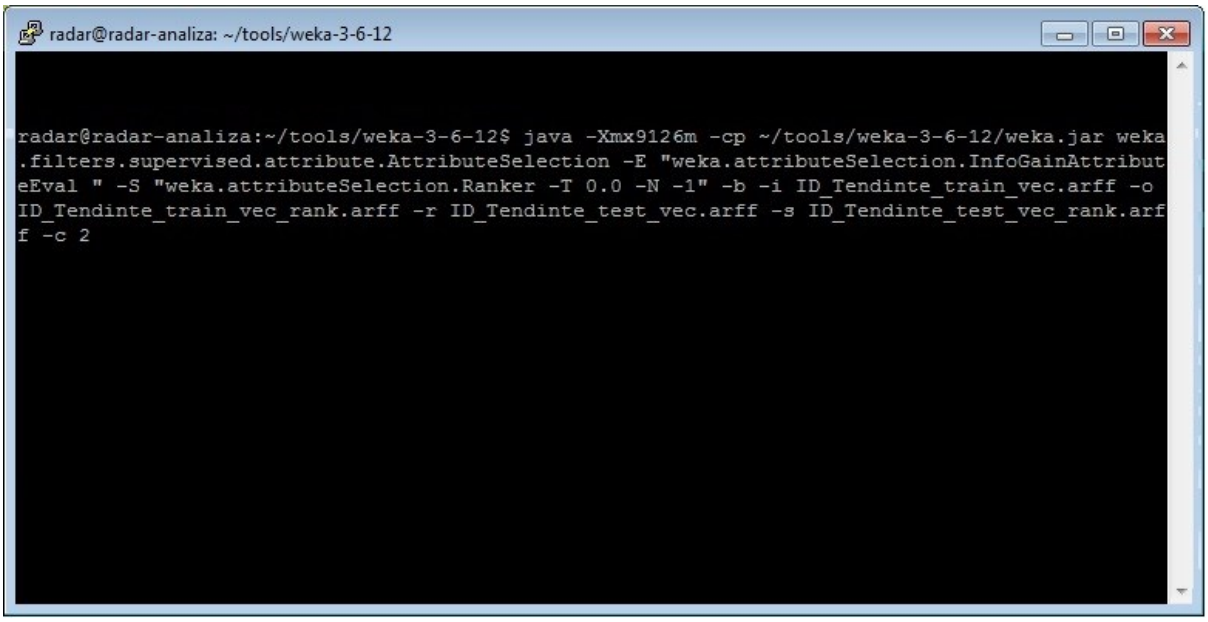


Fig. 25

4. Urmatorul pas este sa folosim functia **weka.filters.supervised.attribute.AttributeSelection,** care selecteaza cele mai importante atribute. Este o etapa care dureaza destul de mult, dar ofera rezultate foarte bune.
5. Dupa alegerea corecta a parametrilor clasificatorului SVM, se va crea un model pe baza textelor folosite pentru invatare. Pentru verificarea acuratetei modelului se va folosi cross-validation. Cross-validarea este o tehnica de validare pentru a evalua modul in care rezultatele unei analize statistice se va generaliza la un set de date independent.
   In acest caz acuratetea modelului este de 93,77%

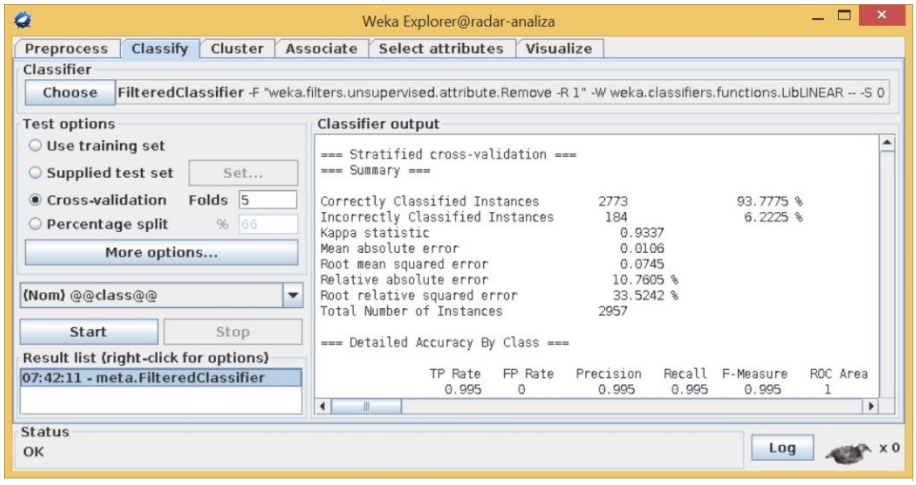


Fig. 26

6. Dupa gasirea celor mai buni parametrii si construirea modelului final aplicam acelasi clasificator pentru corpusul de test, in cazul nostrum– Liblinear

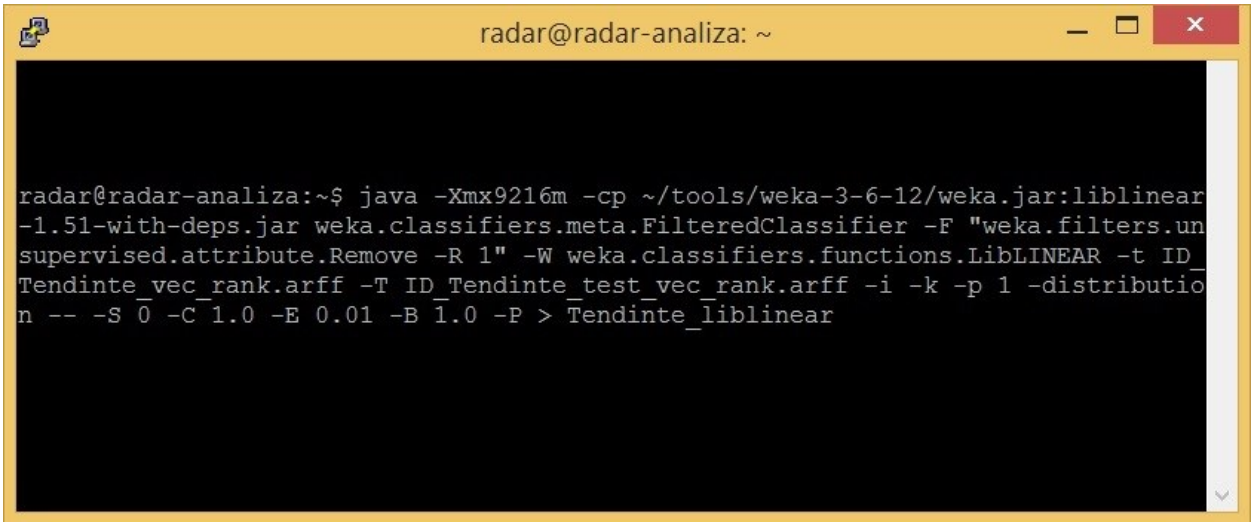


Fig. 27

Obtinanad clasificarea pe categorii a fiecarei stiri in formatul urmator

7. Pentru a da o forma finala a rezultatelor vom folosi un program “perl” facut special pentru acest tip de output.

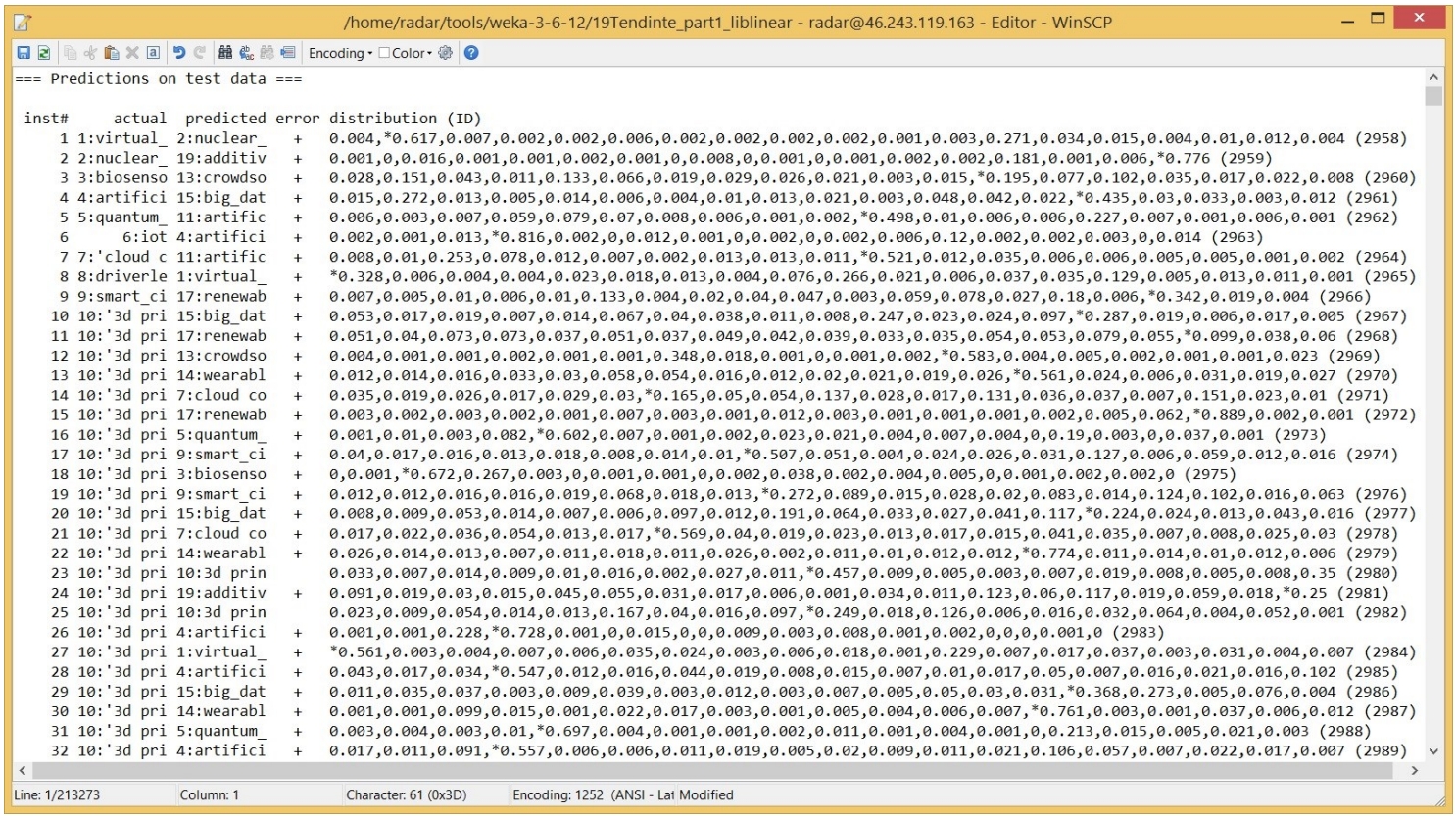

/home/radar/tools/weka-3-6-12/19Tendinte_part1_liblinear - radar@46.243.119.163 - Editor - WinSCP
Encoding
Color
=== Predictions on test data ===
inst# actual predicted error distribution (ID)
1 1:virtual_ 2:nuclear_ + 0.004,*0.617,0.007,0.002,0.002,0.006,0.002,0.002,0.002,0.002,0.001,0.003,0.271,0.034,0.015,0.004,0.01,0.012,0.004 (2958)
2 2:nuclear_ 19:additiv + 0.001,0,0.016,0.001,0.001,0.002,0.001,0,0.008,0,0.001,0,0.001,0.002,0.002,0.181,0.001,0.006,*0.776 (2959)
3 3:biosenso 13:crowdso + 0.028,0.151,0.043,0.011,0.133,0.066,0.019,0.029,0.026,0.021,0.003,0.015,*0.195,0.077,0.102,0.035,0.017,0.022,0.008 (2960)
4 4:artifici 15:big_dat + 0.015,0.272,0.013,0.005,0.014,0.006,0.004,0.01,0.013,0.021,0.003,0.048,0.042,0.022,*0.435,0.03,0.033,0.003,0.012 (2961)
5 5:quantum_ 11:artific + 0.006,0.003,0.007,0.059,0.079,0.07,0.008,0.006,0.001,0.002,*0.498,0.01,0.006,0.006,0.227,0.007,0.001,0.006,0.001 (2962)
6 6:iot 4:artifici + 0.002,0.001,0.013,*0.816,0.002,0,0.012,0.001,0,0.002,0,0.002,0.006,0.12,0.002,0.002,0.003,0,0.014 (2963)
7 7:'cloud c 11:artific + 0.008,0.01,0.253,0.078,0.012,0.007,0.002,0.013,0.013,0.011,*0.521,0.012,0.035,0.006,0.006,0.005,0.005,0.001,0.002 (2964)
8 8:driverle 1:virtual_ + *0.328,0.006,0.004,0.004,0.023,0.018,0.013,0.004,0.076,0.266,0.021,0.006,0.037,0.035,0.129,0.005,0.013,0.011,0.001 (2965)
9 9:smart_ci 17:renewab + 0.007,0.005,0.01,0.006,0.01,0.133,0.004,0.02,0.04,0.047,0.003,0.059,0.078,0.027,0.18,0.006,*0.342,0.019,0.004 (2966)
10 10:'3d pri 15:big_dat + 0.053,0.017,0.019,0.007,0.014,0.067,0.04,0.038,0.011,0.008,0.247,0.023,0.024,0.097,*0.287,0.019,0.006,0.017,0.005 (2967)
11 10:'3d pri 17:renewab + 0.051,0.04,0.073,0.073,0.037,0.051,0.037,0.049,0.042,0.039,0.033,0.035,0.054,0.053,0.079,0.055,*0.099,0.038,0.06 (2968)
12 10:'3d pri 13:crowdso + 0.004,0.001,0.001,0.002,0.001,0.001,0.348,0.018,0.001,0,0.001,0.002,*0.583,0.004,0.005,0.002,0.001,0.001,0.023 (2969)
13 10:'3d pri 14:wearabl + 0.012,0.014,0.016,0.033,0.03,0.058,0.054,0.016,0.012,0.02,0.021,0.019,0.026,*0.561,0.024,0.006,0.031,0.019,0.027 (2970)
14 10:'3d pri 7:cloud co + 0.035,0.019,0.026,0.017,0.029,0.03,*0.165,0.05,0.054,0.137,0.028,0.017,0.131,0.036,0.037,0.007,0.151,0.023,0.01 (2971)
15 10:'3d pri 17:renewab + 0.003,0.002,0.003,0.002,0.001,0.007,0.003,0.001,0.012,0.003,0.001,0.001,0.001,0.002,0.005,0.062,*0.889,0.002,0.001 (2972)
16 10:'3d pri 5:quantum_ + 0.001,0.01,0.003,0.082,*0.602,0.007,0.001,0.002,0.023,0.021,0.004,0.007,0.004,0,0.19,0.003,0,0.037,0.001 (2973)
17 10:'3d pri 9:smart_ci + 0.04,0.017,0.016,0.013,0.018,0.008,0.014,0.01,*0.507,0.051,0.004,0.024,0.026,0.031,0.127,0.006,0.059,0.012,0.016 (2974)
18 10:'3d pri 3:biosenso + 0,0.001,*0.672,0.267,0.003,0,0.001,0.001,0,0.002,0.038,0.002,0.004,0.005,0,0.001,0.002,0.002,0 (2975)
19 10:'3d pri 9:smart_ci + 0.012,0.012,0.016,0.016,0.019,0.068,0.018,0.013,*0.272,0.089,0.015,0.028,0.02,0.083,0.014,0.124,0.102,0.016,0.063 (2976)
20 10:'3d pri 15:big_dat + 0.008,0.009,0.053,0.014,0.007,0.006,0.097,0.012,0.191,0.064,0.033,0.027,0.041,0.117,*0.224,0.024,0.013,0.043,0.016 (2977)
21 10:'3d pri 7:cloud co + 0.017,0.022,0.036,0.054,0.013,0.017,*0.569,0.04,0.019,0.023,0.013,0.017,0.015,0.041,0.035,0.007,0.008,0.025,0.03 (2978)
22 10:'3d pri 14:wearabl + 0.026,0.014,0.013,0.007,0.011,0.018,0.011,0.026,0.002,0.011,0.01,0.012,0.012,*0.774,0.011,0.014,0.01,0.012,0.006 (2979)
23 10:'3d pri 10:3d prin 0.033,0.007,0.014,0.009,0.01,0.016,0.002,0.027,0.011,*0.457,0.009,0.005,0.003,0.007,0.019,0.008,0.005,0.008,0.35 (2980)
24 10:'3d pri 19:additiv + 0.091,0.019,0.03,0.015,0.045,0.055,0.031,0.017,0.006,0.001,0.034,0.011,0.123,0.06,0.117,0.019,0.059,0.018,*0.25 (2981)
25 10:'3d pri 10:3d prin 0.023,0.009,0.054,0.014,0.013,0.167,0.04,0.016,0.097,*0.249,0.018,0.126,0.006,0.016,0.032,0.064,0.004,0.052,0.001 (2982)
26 10:'3d pri 4:artifici + 0.001,0.001,0.228,*0.728,0.001,0,0.015,0,0,0.009,0.003,0.008,0.001,0.002,0,0,0,0.001,0 (2983)
27 10:'3d pri 1:virtual_ + *0.561,0.003,0.004,0.007,0.006,0.035,0.024,0.003,0.006,0.018,0.001,0.229,0.007,0.017,0.037,0.003,0.031,0.004,0.007 (2984)
28 10:'3d pri 4:artifici + 0.043,0.017,0.034,*0.547,0.012,0.016,0.044,0.019,0.008,0.015,0.007,0.01,0.017,0.05,0.007,0.016,0.021,0.016,0.102 (2985)
29 10:'3d pri 15:big_dat + 0.011,0.035,0.037,0.003,0.009,0.039,0.003,0.012,0.003,0.007,0.005,0.05,0.03,0.031,*0.368,0.273,0.005,0.076,0.004 (2986)
30 10:'3d pri 14:wearabl + 0.001,0.001,0.099,0.015,0.001,0.022,0.017,0.003,0.001,0.005,0.004,0.006,0.007,*0.761,0.003,0.001,0.037,0.006,0.012 (2987)
31 10:'3d pri 5:quantum_ + 0.003,0.004,0.003,0.01,*0.697,0.004,0.001,0.001,0.002,0.011,0.001,0.004,0.001,0,0.213,0.015,0.005,0.021,0.003 (2988)
32 10:'3d pri 4:artifici + 0.017,0.011,0.091,*0.557,0.006,0.006,0.011,0.019,0.005,0.02,0.009,0.011,0.021,0.106,0.057,0.007,0.022,0.017,0.007 (2989)
Line: 1/213273
Column: 1
Character: 61 (0x3D)
Encoding: 1252 (ANSI - Lat
Modified

Fig. 28

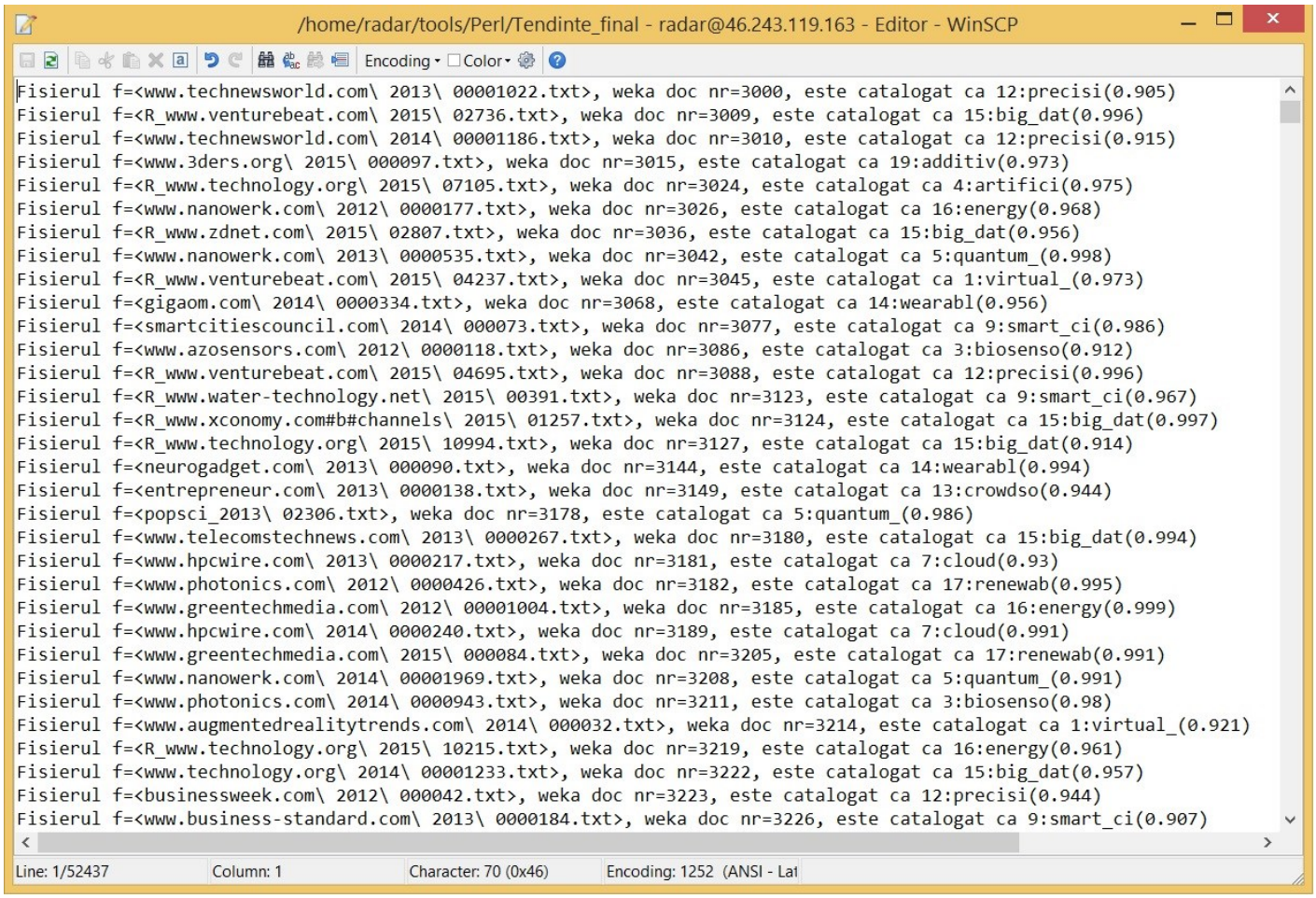


Fig. 29

Fisierul **www.technewsworld.com 2013 00001022.txt** este catalogat ca 12:precision – prescurtare pentru categoria **precision_agriculture** cu o precizie de 90,5%

Proiect cofinanţat din Fondul Social European, prin Programul Operaţional "Dezvoltarea Capacităţii Administrative", în perioada 2007-2013

Listă Anexe:

Lista Anexe:

Anexa 1_Repository _Tagy
Anexa 2_Rezultate surse adiționale știri_similitudine
Anexa 3_Rapoarte_PD_SS_iunie 2015.
Anexa 4_Platforma Tagy
Anexa 5_Raport general NS
Anexa 6_Raport Vot Platforme
Anexa 7_Statistici si analize.actualizare
Anexa 8 Domenii descriptori corpus_final
Anexa 9 SmartDictionary_final
Anexa 10.Dictionary tendinte

Listă Figuri: