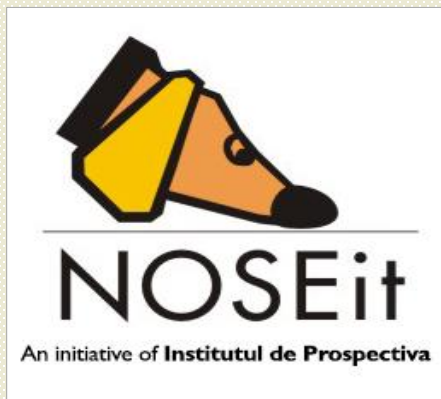




Mecanism de scanare a semnalelor slabe & tendintelor

Echipa: Adrian Curaj, Alina Irimia, Bianca Dragomir, Corina Stefan, Liviu Andreescu, Octavian Popa, Paul Plescan, Paul Punguta, Radu Gheorghiu

Expert extern: Octavian Popescu

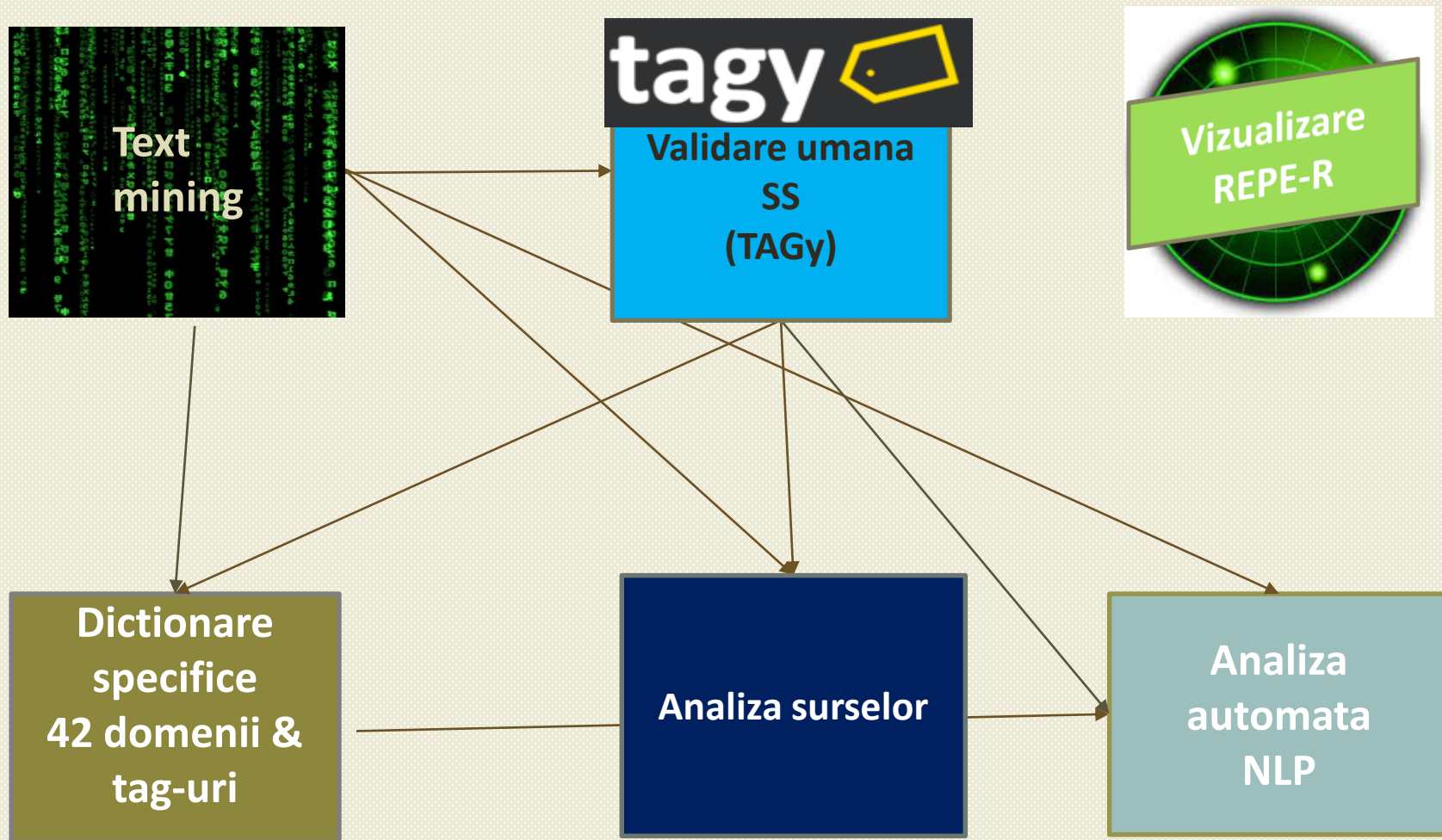
NOSEit – la ce e util un astfel de mecanism?

- Orientare strategica (mediul privat/public)
- Politici bazate pe evidente (evidente – evolutii probabile ale unor fenomene emergente)
- Consultari multi-actor/stakeholder

NOSEit –Componente majore

- Platforma Radar Repository (Corpusuri de stiri)
- Procesul de validare umana(gaming - TAGy) – detectarea de “semnale slabe”
- Modul analiza automata (Natural Language Processing)
- Platforma de vizualizare Repe-R
- Platforma de Facebook NOSEit

NOSEit –componente majore



Repository

O baza de date ce contine ~ 500 000 de stiri din aprox 300 surse online.

20 000 stiri adaugate lunar (*phys.org, nature.com, bbc.com, biosciencetechnology.com, etc.*)

- Preluare automata a stirilor din RSS feed
- Eliminarea duplicatelor (software intern + Plagiarism software)
- Calitatea surselor in termeni de SS => top 30 de platforme de stiri
- Filtre multiple:
 - Sursa
 - Data (zi/luna/an)
 - Tags (SS, domenii)
 - Selectie stiri TAGy

Repository

radarrepository.uefiscdi.ro/Forms/ListOfNews.aspx

Radar Repository Home Administrare » Lista de stiri Bun venit, admin [Delegare]

Lista de stiri

FILTRARE

Text: **Filtering**

Platforma:

Numar intre: -

Tags

Links

	Tags	Link	ACTION
☐	ICT - Internet of Things	http://asmarplanet.com/blog/2015/02/keeping-pace-rapid-rate-big-data-innovation.html	
☐	ICT - Internet of Things	http://asmarplanet.com/blog/2012/09/innovation-or-investments-what-can-transform-the-cities-in-india.html	
☐	ICT - Internet of Things	http://asmarplanet.com/blog/2012/06/innovation-in-motion-enhancing-the-fan-experience-at-the-us-open.html	
☐	ICT - Internet of Things	http://asmarplanet.com/blog/2014/07/can-tech-innovation-help-end-cancer.html	

Platforms

- huffingtonpost.com
- inside3dprinting.com
- earthtechling.com
- businessweek.com
- articles.chicagotribune.com
- environmentalleader.com
- eu-smartcities.eu
- itechfuture.com
- japandailynews.com
- insidehpc.com
- blogs.wsj.com
- lifescience.net
- lifescienceinvestingnews.com
- economist.com
- asmarplanet.com
- greenwisebusiness.co.uk
- campustechnology.com
- insidehighered.com

Repository

Source info

ajnabili.com 1 - Should You Buy a Book Color?

Platforma: ajnabili.com
Nr. ordine: 1
Link Rader: <http://ajnabili.com/tech/should-you-buy-a-nook-color.html>
Data Rader: 2011-12-28
Link original:
Data sura:
Tags import: LIFE SCIENCES / BIOTECH
Fisier: ajnabili.com 2011 00001 [Descarca fisier]

Short text

Text scourt:
Technology has changed the way we do just about everything including how we read books. With an eReader you can read almost any book you want anytime anywhere. No more travelling all the way to the bookstore to pick up a copy. No more trying to find extra space on your shelves to store all [...]

More info

TEXT INTEGRAL

Technology has changed the way we do just about everything including how we read books. With an eReader you can read almost any book you want anytime anywhere. No more travelling all the way to the bookstore to pick up a copy. No more trying to find extra space on your shelves to store all those books you just can't get rid of. No more weighing down your backpack or luggage with reading material. No more paper cuts. EReaders are similar in design to tablet computers but they're less expensive lighter and don't do much else other than their main purpose & allow you to read books. In the beginning of eReaders you could only read books in black and white. In most cases people aren't bothered by the lack of color even if they could imagine the benefits of it because they spend the majority of their time reading simple black text. Traditional eReaders are simplistic. The Nook Color changed things when it was introduced as a color eReader. But if you're going to purchase an eReader do you really need anything more than simple? In other words should you buy a Nook Color? Color vs. Black and White The benefits of color for an eReader are plentiful. It totally enhances the way you can view your magazine subscriptions photographs illustrations children's books and more. The Nook Color makes the reading experience of these types of materials much more enjoyable than other eReaders. Traditional eReaders however use e-ink technology which is a unique technology developed to make reading with it extremely similar to reading actual books. Color e-ink technology does not yet exist for consumers in the US. Therefore the Nook Color is backlit and does not use the same technology as black and white eReaders. The backlighting is good because it makes it possible to read in the dark. However it can strain your eyes or cause headaches after long-term viewing similar to a computer monitor. This doesn't happen with the e-ink used for black and white eReaders. You'll have to weigh the balance & is the ability to view color more important or necessary for you than the ability to read for hours and hours with no eye strain? If so you should purchase a Nook Color. If not you shouldn't. More Functions An additional benefit of the Nook Color is the ability to run thousands of Nook apps. There are apps for just about everything you could need without the overwhelming selection of Apple or Android apps. It also plays music. If you want your eReader to do a lot more than just read books the Nook Color might be for you. However the Nook Color is still limited in what it can accomplish and if you have a tablet already there's probably no real need for the additional functions other than entertainment. There is also no 3G available so if you don't have access to Wi-Fi you can't use most of those additional

Full text

REPE-R

Interfata web de vizualizare a stirilor si incadrare pe domenii

Software-uri utilizate pentru input:

- **Tropes** – software analiza semantica - folosit pentru dezvoltarea dictionarelor pe domenii
- **Zoom** – extensia Tropes, export de stiri in urma indexarii stirilor pe baza unui dictionar
- **OwlEdge** – software de analiza semantica, care poate realiza exporturi xml pe baza dictionarelor realizate in Tropes si poate oferi si alte tipuri de informatii/analize pe baza textelor si a dictionarului.
- **Plagiarism Detector (PD)** = software de detectie a plagiatului/similaritatii. Oferă rapoarte html cu surse aditionale pentru stiri/SS: max 10 surse cu min 50% similaritate pt fiecare stire/ss
- **xPlg** =Raport xls cu sursele indentificare prin PD + date asociate lor.
- **exportBulk** – transforma stirile din excel format txt

REPE-R

POWERED BY *Uesfiscdi* UDI Drive :: Vizualizari :: Ontologie ::

UDI Drive
Vizualizari
Ontologie
Terms and Conditions
Contact
Administrare

Elipsa Stanga
Ramura Stanga
franta Stanga

Background
Resetareza Culoare
Calculeaza intersectia

Elipsa Dreapta
Ramura Dreapta
franta Dreapta

Functionalitati:
- butoane: controleaza culorile elementelor
- mouse scroll: zoom in/out
- click stanga (hold): drag

Nr. crt.
Document

1
Nearly two years after the conclusion of its space shuttle program left Americans wondering what would become of the spacefaring dreams of decades past NASA has sought the advice of health and ethics experts for protecting astronauts on its next generation...

2
ScienceDaily_2014 17805.txt

3
Three-dimensional printers are transforming the business medical and consumer landscape by creating a vast variety of objects including airplane parts football cleats lamps jewelry and even artificial human bones Now astronomers Carol Christian and Antonel...

Panou Stanga

Panou Dreapta

Buton de export a rezultatelor in baza unor filtrari initiale.

Administrare

Zona pentru formule si
diverse queries
Selectie dictionare/import
dictionare (daca vrem sa vedem
doar tendinte de ex)

Filtre pentru: vizualizare grafica (culori care sa
diferentieze diverse topicuri), ordonare/sortare
lista, adaugare frecvente.

Mecanismul de validare umana

Ce ? Clasificarea si incadrarea stirilor in categoria SS sau NS (7 categorii de SS, NS, discard);

Cine? 20 studenti masteranzi(specializari diverse)

- recrutati pe baza abilitatilor de intelegere a textelor in limba engleza;
- organizati in echipe de cate 2 dar care isi schimba compozitia la fiecare flux de lucru;

Cum? Interactiunea se realizeaza online, fiecare actionand atat in calitate de Extractor(E) cat si de Checker(C);

Mecanismul de validare

Cum se intampla?

Etapele validarii:

1. E: extrage ideea de baza din textul integral, apoi incadreaza textul ca SS/NS/discard;
2. C: voteaza textul ca SS sau NS; Daca votul C este identic cu E runda se incheie; daca nu se intampla asa, C poate accepta categoria E sau isi poate pastra opinia prin argumentare;
3. E: Fie accepta argumentul C si alternativa propusa de el sau isi mentine votul original furnizand un ultim argument pentru decizia sa si astfel runda se incheie;
4. Din total SS → Top SS (~ 30/ luna)

Extragere continut si clasificare

Extractor - Extragere continut si clasificare (Finalizata)

- China Combat Smog with Artificial Rain [\[Link\]](#)

China plans to open the heavens to bring back its blue skies. According to local news reports, a document released by the China Meteorological Administration says that from 2015, local weather authorities will be allowed to use cloud seeding to create rain and clear the country's notorious smog. It's part of the government's plan to invest 1.7 trillion yuan (\$277 billion) in tackling air pollution. But will it work? Programmes to generate artificial rain have been running since the 1950s. In recent years, the controversial technique has gained some credibility. "There is indeed something to be said in favour of cloud seeding, provided it is very targeted," says Roger Stone, a climate scientist from the University of Southern Queensland in Toowoomba, Australia. China has used cloud seeding before on a sporadic basis, for example, to ensure clear skies for the opening ceremony of the Beijing Olympics. The technique the Chinese use involves firing rockets carrying a payload of silver iodide particles into the clouds. Ice crystals then form from super-cooled liquid water in the clouds, using the particles as nucleation points. The crystals fall as rain or snow depending on the temperature below. The idea is that the artificial precipitation should clear the smog below. There are other methods that deploy different chemicals designed to work in warmer clouds, but the "glaciogenic" method the Chinese use is considered the most effective, says Steven Siems from Monash University in Melbourne, Australia. However, he doubts cloud seeding will be effective in smoggy conditions. This is because smog is made of particulates of soot and dust, which should already have acted as ice nuclei. If it hasn't rained already, Siems reasons, then adding more particulates in the shape of silver iodide won't help. Satellite data suggest China doesn't have the right atmospheric conditions for cloud seeding, Siems says. "I can pretty much confidently say they don't." And he adds that the skies over China already contain a lot of ice nuclei from factory pollution and dust from the Gobi desert.

Textul extras:

pollution. But will it work? Programmes to generate artificial rain have been running since the 1950s. In recent years, the controversial technique has gained some credibility. "There is indeed something to be said in favour of cloud seeding, provided it is very targeted," says Roger Stone, a climate scientist from the University of Southern Queensland in Toowoomba, Australia. China has used cloud seeding before on a sporadic basis, for example, to ensure clear skies for the opening ceremony of the Beijing Olympics. The technique the Chinese use involves firing rockets carrying a payload of silver iodide particles into the clouds. Ice crystals then form from super-cooled liquid water in the clouds, using the particles as nucleation points. Th

Taxonomie

Categorii alese de dvs.:

- ☒ Innovation in Services
- ☐ Trend shift
- ☒ Behavioral Change
- ☐ Major actor move
- ☐ Discovery
- ☐ Studies
- ☐ Wildcard
- ☐ NS
- ☐ Categorie noua de test
- ☐ Technology

Categorii alese de checker:

- ☒ Innovation in Services
- ☐ Trend shift
- ☒ Behavioral Change
- ☐ Major actor move
- ☐ Discovery
- ☐ Studies
- ☐ Wildcard
- ☐ NS
- ☐ Categorie noua de test
- ☐ Technology

Inchide

Statistici

Statistici pe liste de stiri

Statistici pe liste de stiri

Lista de stiri: **Lista import 1**

Lista import 1

Lista 2

Lista 3

Lista 4

Lista 3 Test Multiplicat

Lista 8_15 decembrie2014

Lista 5 Test

lista 6_2400

lista 23dec_5 ian 2015

Lista 7_11 Ianuarie 2015

Lista 12_18 ianuarie 2015

Lista 19_25 ianuarie 2015

Lista 26.01_01.02.2015

Lista 2_8 Februarie 2015

Lista 3_8 Februarie 2015

Lista 9-15 Februarie 2015

Lista 16-22 Februarie 2015

Lista 23 Februarie-1 Martie 2015

Lista 2_8 Martie 2015

Lista 9_15 Martie 2015

Lista 16_22 Martie 2015

Lista 23_29.03.2015

Status: **- Oricare -**

- Oricare -

Trebuie extrase

In lucru la checker

In lucru la extractor (reverificare)

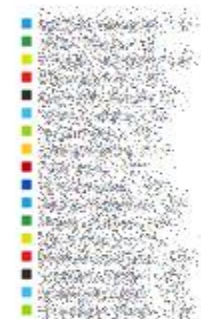
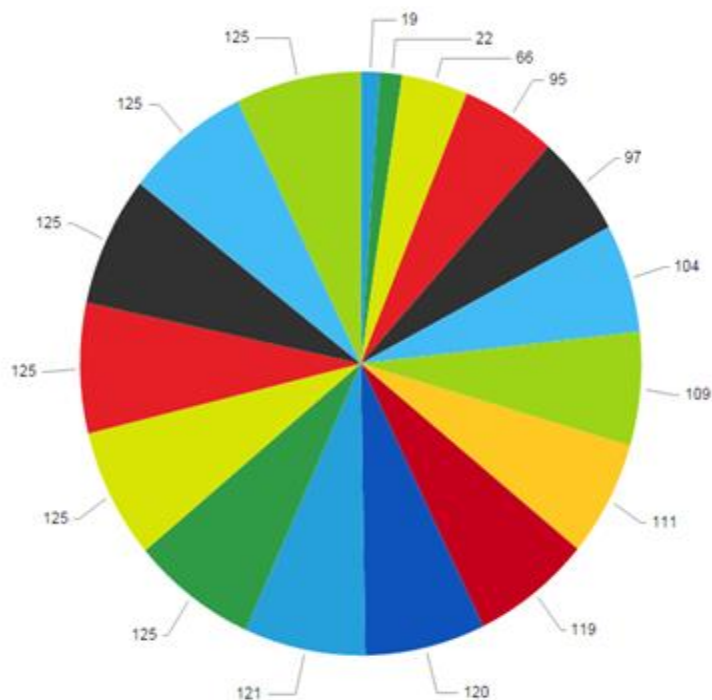
Finalizate

Statistici pe liste de stiri

Statistici pe liste de stiri

Lista de stiri: **Lista 23_29.03.2015** Status: **Trebuie extrase** Afiseaza statisticile

Lista extractori care mai au stiri de extras



Rezultate: ~ 5000 SS.

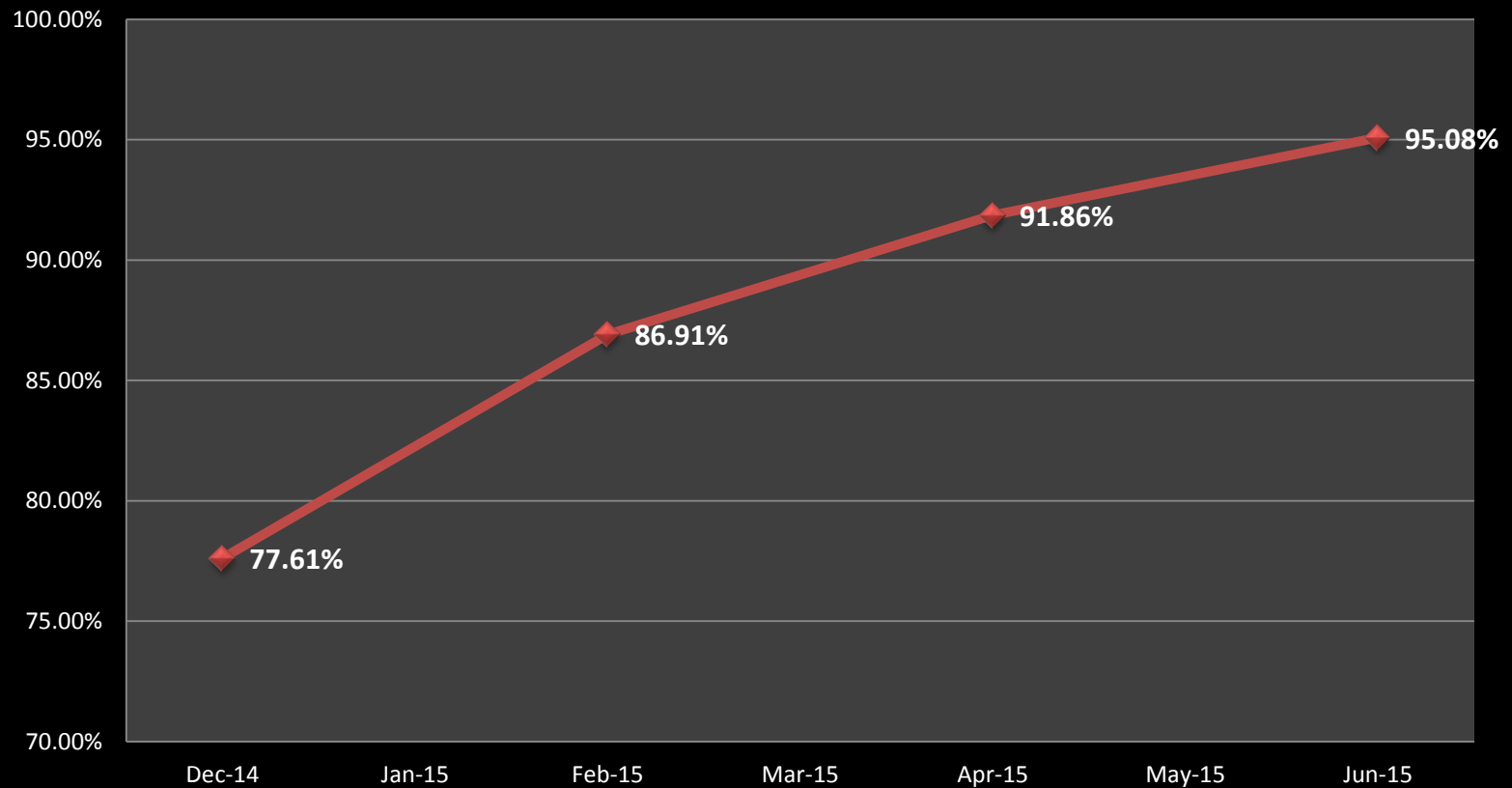
Dec 2014 – Iunie 2015	Nr. stiri	SS	NS + Discarded	SS %
Seturi de stiri 1-4	6800	1364	5436	20.06%
Seturi de stiri 5-6	4000	402	3598	10.05%
Seturi de stiri 7-13	15875	617	15258	3.89%
Total	26675	2383	24292	8.93%
Seturi de stiri 14-27	43045	1454	41591	3.37%
Total	69720	3837	65883	5.5%

Signal categories	Nr. voturi	%
<i>Technology</i>	2309	3.31 %
Behavioral change	77	0.11 %
Innovation in services	228	0.33 %
Major actor move	318	0.46 %
Trend shift	186	0.27 %
Discovery	823	1.18 %
Wildcard	74	0.11 %
NS	62882	90.19 %
<i>Discarded</i>	2996	4.30 %

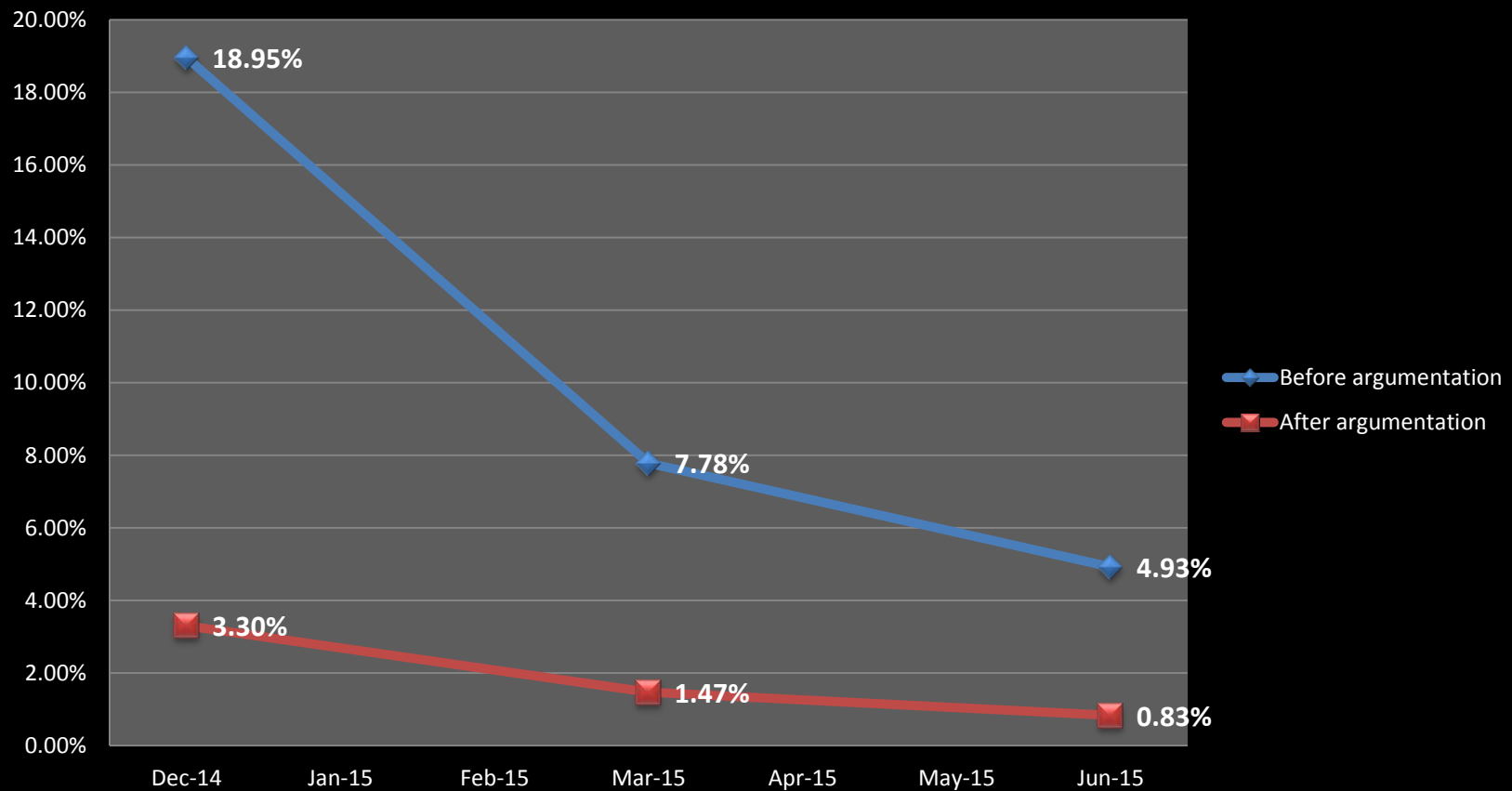
Evolutia consensului

Predictii bune= Atat E cat si C voteaza identic de la inceput

Evolutia predictiilor bune



Reducerea divergentei in timp



Modul NLP - Natural language processing

Utilizarea unor algoritmi de inteligența artificială:

- SVM - tehnica Support Vector Machine permite ca, pornind de la un set de texte exemplu dintr-un domeniu (ex. foraj marin), să selecteze din corpusul de știri pe cele care se încadrează în acest domeniu (știrile sunt ordonate în funcție de probabilitatea de încadrare).
- LDA - Latent Dirichlet Allocation e metoda de procesare a limbajului care permite clasificarea semantică a unui corpus mare de știri într-un număr de clustere la alegerea analistului uman;
- Algoritmi de machine learning care permit învățarea unor modele/patternuri semantice din textele validate ca SS și detectarea lor în alte texte, cu anumită probabilitate

<https://www.facebook.com/NOSEit>

