

Opinion Mining on News Portal Comments – Towards a Croatian Sentiment Database –



How it feels

To disagree with the general opinion on an internet forum

Markus Schatten, PhD

University of Zagreb
Faculty of Organization and Informatics

May 8, 2012

Introduction

Challenges

Challenges

Opinion Repositories

Methods

News Portals

Case Study - Croatian News Portals

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- “What other people think” has always been an important piece of information for most of us during the decision-making process.

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- “What other people think” has always been an important piece of information for most of us during the decision-making process.
- What can be done with subjective information:

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- “What other people think” has always been an important piece of information for most of us during the decision-making process.
- What can be done with subjective information:
 - Product/services research for marketing purposes

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- “What other people think” has always been an important piece of information for most of us during the decision-making process.
- What can be done with subjective information:
 - Product/services research for marketing purposes
 - Better search engines

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- “What other people think” has always been an important piece of information for most of us during the decision-making process.
- What can be done with subjective information:
 - Product/services research for marketing purposes
 - Better search engines
 - Campaign monitoring (e.g. political, business, showbiz, ...)

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- “What other people think” has always been an important piece of information for most of us during the decision-making process.
- What can be done with subjective information:
 - Product/services research for marketing purposes
 - Better search engines
 - Campaign monitoring (e.g. political, business, showbiz, ...)
 - Recommendation systems

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- “What other people think” has always been an important piece of information for most of us during the decision-making process.
- What can be done with subjective information:
 - Product/services research for marketing purposes
 - Better search engines
 - Campaign monitoring (e.g. political, business, showbiz, ...)
 - Recommendation systems
 - Detection of “flames”

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- “What other people think” has always been an important piece of information for most of us during the decision-making process.
- What can be done with subjective information:
 - Product/services research for marketing purposes
 - Better search engines
 - Campaign monitoring (e.g. political, business, showbiz, ...)
 - Recommendation systems
 - Detection of “flames”
 - Public health (e.g. emotional/mental status, detecting potential suicide victims ...)

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- “What other people think” has always been an important piece of information for most of us during the decision-making process.
- What can be done with subjective information:
 - Product/services research for marketing purposes
 - Better search engines
 - Campaign monitoring (e.g. political, business, showbiz, ...)
 - Recommendation systems
 - Detection of “flames”
 - Public health (e.g. emotional/mental status, detecting potential suicide victims ...)
 - Government intelligence

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- “What other people think” has always been an important piece of information for most of us during the decision-making process.
- What can be done with subjective information:
 - Product/services research for marketing purposes
 - Better search engines
 - Campaign monitoring (e.g. political, business, showbiz, ...)
 - Recommendation systems
 - Detection of “flames”
 - Public health (e.g. emotional/mental status, detecting potential suicide victims ...)
 - Government intelligence
 - Trend monitoring (e.g. social, cultural, ...)

Introduction

- Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.
- “What other people think” has always been an important piece of information for most of us during the decision-making process.
- What can be done with subjective information:
 - Product/services research for marketing purposes
 - Better search engines
 - Campaign monitoring (e.g. political, business, showbiz, ...)
 - Recommendation systems
 - Detection of “flames”
 - Public health (e.g. emotional/mental status, detecting potential suicide victims ...)
 - Government intelligence
 - Trend monitoring (e.g. social, cultural, ...)

Challenges

- Contrasts with Standard Fact-Based Textual Analysis
 - Categorization vs. Continuum
 - Taxonomy vs. Intensity of feelings

Challenges

- Sentiment polarity

Challenges

- Sentiment polarity (Pang et al., 2002)

Challenges

- Sentiment polarity (Pang et al., 2002)

	Proposed word list	Accuracy
Human 1	positive: dazzling, brilliant, phenomenal, excellent, fantastic negative: suck, terrible, awful, unwatchable, hideous	58 %
Human 2	positive: gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting negative: bad, cliched, sucks, boring, stupid, slow	64 %

Challenges

- Sentiment polarity (Pang et al., 2002)

	Proposed word list	Accuracy
Human 1	positive: dazzling, brilliant, phenomenal, excellent, fantastic negative: suck, terrible, awful, unwatchable, hideous	58 %
Human 2	positive: gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting negative: bad, cliched, sucks, boring, stupid, slow	64 %
Statistics-based	positive: love, wonderful, best, great, superb, still, beautiful negative: bad, worst, stupid, waste, boring, ?, !	69 %

Challenges

- Domain dependent

Challenges



- Domain dependent
 - The same sentence can mean two very different things in different domains

Challenges

- Domain dependent
 - The same sentence can mean two very different things in different domains
 - “Read the book.” ← Good for books, bad for movies

Challenges

- Domain dependent
 - The same sentence can mean two very different things in different domains
 - “Read the book.” ← Good for books, bad for movies
 - “You’re in for one hell of a bumpy ride!” ← Good for movies and books, bad for cars.

Challenges

- Domain dependent
 - The same sentence can mean two very different things in different domains
 - “Read the book.” ← Good for books, bad for movies
 - “You’re in for one hell of a bumpy ride!” ← Good for movies and books, bad for cars.
 - Sentimental word associations change with domain

Challenges

- Domain dependent
 - The same sentence can mean two very different things in different domains
 - “Read the book.” ← Good for books, bad for movies
 - “You’re in for one hell of a bumpy ride!” ← Good for movies and books, bad for cars.
 - Sentimental word associations change with domain
 - Fuzzy cameras are bad, but fuzzy teddy bears are good.

Challenges

- Domain dependent
 - The same sentence can mean two very different things in different domains
 - “Read the book.” ← Good for books, bad for movies
 - “You’re in for one hell of a bumpy ride!” ← Good for movies and books, bad for cars.
 - Sentimental word associations change with domain
 - Fuzzy cameras are bad, but fuzzy teddy bears are good.
 - Big trucks are good, but big smart phones are bad.

Challenges

- Domain dependent
 - The same sentence can mean two very different things in different domains
 - “Read the book.” ← Good for books, bad for movies
 - “You’re in for one hell of a bumpy ride!” ← Good for movies and books, bad for cars.
 - Sentimental word associations change with domain
 - Fuzzy cameras are bad, but fuzzy teddy bears are good.
 - Big trucks are good, but big smart phones are bad.
 - Bad is bad, but bad villains are good.

Challenges

- Parody, Satire & Sarcasm



- “I love the cover!” (Book)
- “Be sure you save your purchase receipt.” (Smart phone)
- “Great for insomniacs” (Book)
- “All the features you want. Too bad they don’t work!” (Smart phone)
- “Great idea, now try again with a real product development team” (E-reader)

Challenges

- Thwarted expressions

Challenges



- Thwarted expressions
 - The sentences/words that contradict the overall sentiment of the comment are in majority

Challenges

- Thwarted expressions
 - The sentences/words that contradict the overall sentiment of the comment are in majority

I hate the Spice Girls.... [3 things the author hates about them]... Why I saw this movie is a really, really, really long story, but I did, and one would think I'd despise every minute of it. But... Okay, I'm really ashamed of it, but I enjoyed it. I mean, I admit it's a really awful movie, ... [they] act wacky as hell ... the ninth floor of hell ... a cheap [beep] movie ... The plot is such a mess that it's terrible. But I loved it.

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
- Blogosphere / Microblogging / Vlogs ...

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
- Blogosphere / Microblogging / Vlogs ...
- Forums

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
- Blogosphere / Microblogging / Vlogs ...
- Forums
- Social networking sites

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
- Blogosphere / Microblogging / Vlogs ...
- Forums
- Social networking sites
- Podcasting sites

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
- Blogosphere / Microblogging / Vlogs ...
- Forums
- Social networking sites
- Podcasting sites
- Usenet/Mailing lists/Discussion groups

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
- Blogosphere / Microblogging / Vlogs ...
- Forums
- Social networking sites
- Podcasting sites
- Usenet/Mailing lists/Discussion groups
- ...

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
 - Blogosphere / Microblogging / Vlogs ...
 - Forums
 - Social networking sites
 - Podcasting sites
 - Usenet/Mailing lists/Discussion groups
 - ...
 - Review sites (products, movies, books ..)

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
 - Blogosphere / Microblogging / Vlogs ...
 - Forums
 - Social networking sites
 - Podcasting sites
 - Usenet/Mailing lists/Discussion groups
 - ...
 - Review sites (products, movies, books ..)
 - Rating sites

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
 - Blogosphere / Microblogging / Vlogs ...
 - Forums
 - Social networking sites
 - Podcasting sites
 - Usenet/Mailing lists/Discussion groups
 - ...
 - Review sites (products, movies, books ..)
 - Rating sites
 - Community sites

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
 - Blogosphere / Microblogging / Vlogs ...
 - Forums
 - Social networking sites
 - Podcasting sites
 - Usenet/Mailing lists/Discussion groups
 - ...
 - Review sites (products, movies, books ..)
 - Rating sites
 - Community sites
 - Specialized blogs

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
 - Blogosphere / Microblogging / Vlogs ...
 - Forums
 - Social networking sites
 - Podcasting sites
 - Usenet/Mailing lists/Discussion groups
 - ...
 - Review sites (products, movies, books ..)
 - Rating sites
 - Community sites
 - Specialized blogs
 - News portals with enabled commenting/rating

Opinion Repositories

- With the Social Web explosion there are huge repositories of opinionated text/multimedia available:
 - Blogosphere / Microblogging / Vlogs ...
 - Forums
 - Social networking sites
 - Podcasting sites
 - Usenet/Mailing lists/Discussion groups
 - ...
 - Review sites (products, movies, books ..)
 - Rating sites
 - Community sites
 - Specialized blogs
 - News portals with enabled commenting/rating
 - ...

Blogosphere



- Technorati currently claims it tracks over 113 million blogs

Blogosphere

- Technorati currently claims it tracks over 113 million blogs
- 12,000 blogs created each day

Blogosphere

- Technorati currently claims it tracks over 113 million blogs
- 12,000 blogs created each day
- One new weblog started every 7.4 sec

Blogosphere

- Technorati currently claims it tracks over 113 million blogs
- 12,000 blogs created each day
- One new weblog started every 7.4 sec
- Social networks/blogs are now the fourth most popular on-line category – ahead of e-mail!

Methods

- Machine learning

Methods

- Machine learning
- Natural language processing / computational linguistics

Methods

- Machine learning
- Natural language processing / computational linguistics
- Context mining (social, sentiment holder, domain specifics etc.)

Methods

- Machine learning
- Natural language processing / computational linguistics
- Context mining (social, sentiment holder, domain specifics etc.)
- Summarization techniques (results/knowledge representation)

Methods

- Machine learning
- Natural language processing / computational linguistics
- Context mining (social, sentiment holder, domain specifics etc.)
- Summarization techniques (results/knowledge representation)
- ...

News Portals



The screenshot shows the top navigation bar of the 24 News portal with categories: HOME, NEWS, SHOW, SPORT, LIFESTYLE, and TECH. Below the navigation bar, there are two main news items:

- Vračbini - popusti do 90%**
Kakav je ovaj vrabiti dan? Najbolje nam odgovara meni ekskluzivno poručiti!
Kakavlika.hr/Vrabiti
- Razor Srbije i Kosova**
Razne verzije na glavni Srbije: Čitavi nove vijesti i više foto
www.24Times.com

Below the news items, there is a section for comments:

IZDVOJENI KOMENTARI: 108 komentara [Pogledaj sve komentare](#)

opinijski2: Pa valjda bi Linc prvo trebao provjeriti kako je Todorčić uspio ubao u posao sa svojcem i ojim kapitalom, jerma je tajna vijest države i navetiko se priča da su neistajale ogromne zahtjeve humanitarne pomoći, mnogi ljudi to znaju. Jedn važno je samo tajna ove države je da se na posluim radni i svojim rukama ne može zaraditi sve ono što Todorčić i njegovu slični imaju. Gosp. Linciću, jedno javno pitanje za vas: [pjd... arhivirane](#)

Marko-Burjak: Da je poznata tipica rigne četvrti nikad oni ne bi ni bi lakimiti [pjd... arhivirane](#)

klitka: Poznata im može palnuti u "pakicu" skup u Lisincem, u nakaj državi je optič poznato da se samo "digne proširni", a porije nalazno sve po slatam...

Ne slažeš se? Komentiraj!
Za komentiranje je potrebno prijaviti se. Nemaš korisnički račun? Registracija je

News Portals

24 HOME NEWS SHOW SPORT LIFESTYLE TECH

Veraždin - popusti do 90%
Kakav će biti ovaj dan njegova metu ekološki prihvatljiv?
kolektiva.hr/Veraždin

Razdor Srbije i Kosova
Razne verzije na granici Srbije i Čade
kave sjele i vatre što
www.94Times.com

IZDOVOJEN KOMENTARI: 108 komentara [Pogledaj sve komentare](#)

prije 2 sata 50 minuta
opinijski2: Pa valjda bi Linc prvo trebao provjeriti kako je Todorčić uspio ušao u posao sa cvjećem i ožim kapitalom, javno je tajna cijele države i navetlo se priča da su neke tajne ogromne zaštitne humanitarne pomoći, mnogi ljudi to znaju. Jednako javno tajna ova država je da se na posao radi i svojih rukama ne može zaraditi sve ono što Todorčić i njegovu silnici imaju. Gosp. Linciću javno pitanje za vas puč... [ovakvi sve](#)

prije 4 sata 10 minuta
Markus-Burjak: Da je poznata tipica rigne čudnjak nikad oni ne bi ni bili tajkumiti

prije 4 sata 20 minuta
klitka: Porezna im može pomoći u "pakicu" sklop u Lincićem, u nakoj državi je opće poznato da se samo "digne proširina", a porijeklo nalazno sve po stranom...

Ne slažeš se? Komentiraj!

Za komentiranje je potrebno prijaviti se. Nemaš korisnički račun? [Registracija je](#)

[Facebook](#) (47)

KOMENTARI (36)

Komentari na članak obično se u rednom nizu prikazuju prema broju glasova koji su dobili. Zbog toga je važno biti prvo na listi komentara. Tako se lakše odražavaju na članak, a također i na preglednost i brzinu učitavanja.

Markus-Burjak Otpisano prije sat
Ako ih se skupi više od 150 to će biti mega-skup. Normalno očekuje se gamat sa FF i neradnici iz crvene akcije te još poruci rječnikar i zburjari agraristi

klitka Otpisano prije 1h
Sala.

Markus-Burjak Otpisano prije 1h
eto 12 i globalna rotacija
aH... mesu to opet

News Portals

24 HOME NEWS SHOW SPORT LIFESTYLE TECH

Veraždin - pupasti do 90%
Kolektivni vozi izvan domaćeg mesta
Kolektivna.hr/Veraždin

Razor Srbije i Kosova
Razne verzije na granici Srbije i
Kosova
www.94Times.com

IZDVOJENI KOMENTARI 108 komentara [Pogledaj sve komentare](#)

opširnost: Pa valjda bi Linc prvo trebao provjeriti kako je Todorčić uspio ući u
posao sa cvjećem i Ojnim kapitulacijom, javna je tajna riječi države i navetiko se priča
da su meštaci ogromne zaštitne humanitarne pomoći, mnogi ljudi to znače. Jednako
javnost tajna ove države je da se na poslu radni i svojih rukama ne može zaraditi
sve ono što Todorčić i njegovi učestvovali Gogo Linciću, jedno javno pitanje za vas
p.u., [ovako](#) [ovako](#)

Radnik u Srbiji: da je poznata riječ ipak čudni nikad oni ne bi bi li tajkanti
p.u. 4 [ovako](#) [ovako](#)

klasa: Porezna im može pomoći u "paketu" sklop u Linciću, u nakoj državi je opće
poznato da se samo "digne proširivati", a porjez nastavio sve po starom...

Ne slažeš se? Komentiraj!

Za komentiranje je potrebno prijaviti se. Nemaš korisnički račun? [Registracija](#) je



Više o...
Zoran Mamić

KOMENTARI

26 komentara

[Komentiraj](#)



Saveta Pašković [*](#) Česti komentator - Gimnazija

A gdje su ti veliki poreznici bili sve ove godine pa njih bi trebalo
kazniti da sad oni plate sve taj porez jer su i oni sudjelovali i
pomagali u toj utaji poreza ili je njihov posao bio da nema
jedinicima samo naplaćuju od ove rase bijede ali i sad od togance
bi rasta odim treća se brla rade se mi.
Odgovor [*](#) 30 Svuda mi se - Prati objavu - 16 hours ago



Ivan Kretulović [*](#) Predtjebli se - [*](#) Česti komentator

Svi ti poreznici su primali plaću za rad na nekoje
stranice jače i njegove komunističke partije. Sad samo
da i ovi novi ne provode istu politiku.
Odgovor [*](#) 20 Svuda mi se - 16 hours ago

KOMENTARI (36)

Komentari na članak obično se u redovnom redoslijedu prikazuju prema broju glasova za ili protiv. Zbog toga je
prijedlog: omogućiti izmještanje glasova u određeni redoslijed (na primjer, u redoslijedu najprije najviše glasova).

anaprijet [*](#) Otpovieno prije sat
Ako ih se sklopi više od 150 to će biti megasquash. Normalno očekuje se gamad sa FF i
neradnici iz crvene akcije je još porokli rješavanje i zburjenje agnosti

halibit [*](#) Otpovieno prije 1h
Sala.

stargoma [*](#) Otpovieno prije 1h
eto 12 og globalna rotacijska
aH... neka to opet

News Portals



24 HOME NEWS SHOW SPORT LIFESTYLE TECH

Varaždin - popusti do 90%
Kolektiva vam iznosi novi odgovorni menu ekskluzivno ponudu!
Kolektiva.hr/Varaždin

Razor Srbije i Kosova
Razne verzije na granici Srbije, Crne Gore i Srbije i Kosova
www.24Times.com

IZDVOJENI KOMENTARI 108 komentara [Pogledaj sve komentare](#)

opinijski2 Pa valjda tu Linc prv trebao provjetiti kako je Todorčić uspio uletao u posao sa cvjetom i Ojini kapitulacijama je lagno obje države i naveli su priča da su nestajale ogromne cefine humanitarne pomoći zbog to znanja Jek valja jama lagno ove države je da se na poslom sudu i englezi rukama ne može zaraditi sve ono što Todorčić i njegovu službu smjaga Gogo Lincija jedino javno pitanje za vas pu... [otkrij sve](#)

RepublikaSRM da je poznata njema rpa čudnjak nikad oni ne bi ni bili tajkanti prije 40 miliona

litka Porezna im može puzati u "paklu" sklop u Lisinsku, u našoj državi je opće poznato da se samo "digne puštinu", a porcije nastavio sve po starom...

Ne slažeš se? Komentiraj!

ZA komentiranje je potrebno prijaviti se. Nemaš komentarni račun? Registracija je

[Komentiraj](#) [Komentiraj](#) (47)

KOMENTARI (36)

Komentari na članak: **Ujedinjeni protiv fašizma** - Nemaš komentarni račun? Registracija je

anapetna Otpisano prije 1st
Ako ih se sklopi više od 150 to će biti megasuegah. Normalno očekuje se garnad sa FF i nekadno iz crvene akcije je još porokri rješenja i džupirni agronosi

halak7 Otpisano prije 1h
tako.

anapetna Otpisano prije 1h
oko 12 og globalna rotaridacija sht... neka to opet

Više o...
Zoran Mamić

KOMENTARI

26 komentara [Komentiraj](#)

Saveta Palkov [Komentiraj](#) **Česti komentator** - Gimnazija
A gdje su ti veliki poreznici bili sve ove godine pa njih bi trebalo kazniti da sad oni plate sva taj porez jer su i oni sudjelovali i pomagali u toj utrazi poreza ili je njihov posao bio da nama jadnicima samo naplaćuju od ove nase bijede ali i sad od togance ne bih htela oim treca se brida rado se mis.

Odgovor [K](#) 30 Svuda mi se - Prati objavu - 16 hours ago

Ivan Krstulović [Komentiraj](#) **Prodjebli se** - **Česti komentator**
Svi ti poreznici su primali plaću za neras po nekog drugarionu jace i ispreke komunističke partije. Sad samo da i ovi novu ne provode isti politiku.

Odgovor [K](#) 20 Svuda mi se - 16 hours ago

Izvor: [Ujedinjeni protiv fašizma](#)

Besplatni internet u RH
Surfaj besplatno mobilnom. Nove Vip Tarife Bez limita.
[www.vip.hr/tafe/nae/india](#)

Google ogasi

Komentiraj

[Odgovori na Facebooku](#)

Vladimir Bačić
Svi Jugoslavani JEDITE GOVNA! I je sam Protiv fašizma! Pa stožo Trina, u HRVATA nije nikada ni bilo NASTA. TO sta Vi Jugoslavani izmislili! Ante PAVELI je na Stranju 26. 1941 Glasavao u HSABORU i od 196 prisutnih, 195 je glasavao PROTIV INŠIZMA! Jebo Ve STITO Židov i Jugoslaven, MASOR, PAVELI i Anti-Hrvati On je Poljsko-SuaKI ŽIDOVNI KNOKI! ŽIDOKI GLUPTI Sve samo Židovi su Pravili i Prvu i Drugu ŽIGU! Kako su radili tako su i prošli. Vi ste Budalni Glupasti! Jebo vas PAKAI! Njemu samo lažne gušteljima! Iživu KRALJEVSTVO HRVATA! Glavni Grad DUVNO! Jebo vas pederski Zagrebi! Hrvatsko Kraljevstvo je: Oš SDO, Rojama, Orina, Drava i (Srijem-Sana), Baščka, Bosna, Crna Gora, Sarčak! Jebo vas Tuđman koji pola Bosne daje Cetinjska! Jebo vas Alpa Turčini Oni porokale, a Bosnom sa šepuri Cetnik Diodi!

Odgovor [K](#) Svuda mi se - Prati objavu - 12 hours ago

[Facebook](#) [Otvoriti u novom prozoru](#)

Komentari

trifone 2	06.05.2012. 15:23
propagata te jumbo je bila hahahahaha	06.05.2012. 15:23
trifone	06.05.2012. 15:23
dormna Otom kak vas u bolja hahahahahah	06.05.2012. 15:24
javeli	06.05.2012. 15:24
lax, pa kaj ne vspati deti svoj rick, ropt...! bi bolji postite roci kak svagdje tebe vidim, strahopadec jet...	06.05.2012. 15:24
trifone	06.05.2012. 15:24
biho je sam priznalo da je dnevinka	06.05.2012. 15:24
javeli	06.05.2012. 15:24
strahopadci... ajde, naravno sad odnava papjeka da ne odnra... a teta rikicu nek usput dnev...	06.05.2012. 15:26
trifone	06.05.2012. 15:26
javeli, Ona me druzba na bati... pozdrav prijatelju, nemam cija hahahahahah... odnvi bezdane vido!	06.05.2012. 15:27
da je prazna porta pa nemaaj kad sa svojom družina hahah... pozdr	06.05.2012. 15:27
javeli	06.05.2012. 15:27

News Portals

24 HOME NEWS SHOW SPORT LIFESTYLE TECH

Varaždin - pupuzi do 90%
Kolektiva vasa izradila novi odgovoran menu...
Kolektiva.hr/Varaždin

Razor Srbije i Kosova
Razor Srbije i Kosova...
www.24Times.com

IZDVOJENI KOMENTARI 108 komentara [Pogledaj sve komentare](#)

opinijski2 Pa valjda to Lincir prvo trebalo provjeriti kako je Todorčić uspio uletao u posao sa cvjećem i Ojnim kapitulacijama je lagna vijest...
MarinaBurmak da je poznata vijest rign...
klita Porezna im može pomoći u "paketu" sklop u Lisinskom, u našoj državi je opće poznato da se samo "digne plaćanje", a porazje nastavio sve po starom...

Ne slažeš se? Komentiraj!
Za komentiranje je potrebno prijaviti se. Nemaš komentarni račun? Registriraj je

KOMENTARI (36)

anjanovic Otpisano prije 1st
Ako ih se slaže više od 150 to će biti megaslagah. Nemašno običuje se gamat sa FF i nekadiz to crvine nekije je još prosaki rjevanac i džupniji agronak

halabica Otpisano prije 1h
tako.

stjepanica Otpisano prije 1h
eto 12 i golobina rotaridacija...
...neca to opet

Više O...
Zoran Mamić

KOMENTARI

26 komentara [Komentiraj](#)

Saveta Palkov Česti komentator - Gimnazija
A gdje su ti veliki poreznici bili sve ove godine pa njih bi trebalo kazniti da sad oni plate nase taj pomez jer su i one sudjelovali i pomagali u toj utaji poreza ili je njihov posao bio da nama javnicima samo naplacuju od ove nase bjeđe ali i sad od togancee bio nista oim treca se brida nadio se nio.
Odgovor: 30 Svuda mi se - Prati objavu - 16 hours ago

Ivan Krutolović Predtjehbi se - Česti komentator
Svi ti poreznici su primali plaću za nerasi po nalozu drugarion jace i ispreke komunističke partije. Sad samo da i ovi nuro ne provode istu politiku.
Odgovor: 20 Svuda mi se - 16 hours ago

Izvor: [Ujedinjeni protiv fašizma](#)

Besplatni internet u RH
Surfaj besplatno mobilnom. Nove Vip Tarife Bez limita.
[www.hrt.hr/nc/nc/10114](#)

Google+ ogasi

Komentiraj

anjanovic Otpisano prije 1h
Odgovori na Facebooku

Vladimir Baotic
Svi Jugoslavieni JEDITE GOVNA! I je sam Protiv fašizma! Pa stoika Trina, u HRVATA nije nikada ni bilo NASTAJA. TO je VI Jugoslavieni izmislili Ante PAVELIC je na Stranice 26. 1941 Glasavao u HŠBORU i od 196 prisutnih, 195 je glasavao PROTIV HRVATA! Jebo VE STITO Židovi i Jugoslaveni, MASOR, PAVELIC i Anti-Hrvati On je Poljako-Buaki ŽIDOVNI KROVI! JEKO GLUPFI Sve samo Židovi su Pravili i Prvu i Drugu JUGU! Kako su radili tako su i prosli. Vi ste Budalni Glupstari! Jebo vas PAVAN! Njemu samo lažne gušteljstini! TRIVU KRALJEVSTVO HRVATA! Glavni Grad DUVNO! Jebo vas pederski Zagrebi! Hrvatsko Kraljevstvo je: Oš SDOZ, Rojama, Orina, Drava i Srijem-Sanati. Bašička, Borna, Crna Gora, Sarčički! Jebo vas Tuđman koji pola Borne daje Cetinica! Jebo vas Alipa Turčin! On porokale, a Bornaom sa šepuri Cetnik Diodaj!
Odgovor: 5 Svuda mi se - Prati objavu - 12 hours ago

Komentari

trifone 2	06.05.2012. 15:23
propisane te Junbo je bio hahahahaha	06.05.2012. 15:23
trifone	06.05.2012. 15:24
domina Otom kak vas u biju hahahahahah	06.05.2012. 15:24
gavel	06.05.2012. 15:24
ta, pa kaj ne vasaš det svoj nica, rovet...! I biš pravije reci kak svagdje tebe vidim, strahopetac jet...	06.05.2012. 15:24
trifone	06.05.2012. 15:24
bišo je sam priznalo da je dnevniha	06.05.2012. 15:24
gavel	06.05.2012. 15:26
strahopetac... ajde, naravi sad odnosa paprika da me odnra... a teta rikicu nek usput dnevni...	06.05.2012. 15:26
trifone	06.05.2012. 15:26
gavel, čeka me dnevno na bati... pozost prijatelj, nemam oca hahahahahah... ostavi bedake višić da je prazna porta pa nemaš kaj sa svojom budnata hahah... pozost	06.05.2012. 15:27

Vaši komentari: (18)

napomena: komentari čitajući ne predstavljaју ni bilo koj maži stav uredništva portala advance.hr. Za sadržaj i točnost komentara čitajući ne odgovaramo. Sadržaj se periodično pregledava i nepravilni komentari se uklanjaju.

gnzonit 01.01.09.05.12

pa pustio nas je 1% da mislimo da smo bili i sad smo slobodni i privatizirani. I na sljedećim izborima mislimo da smo slobodni. Sloboda je sigurnost makar postojе i oprečna misljenja, i to je dobro jer imamo o cemu raspravljajugimertima. A interpunkcija može biti razlog dovoljno ali sam između ostalog postao i bespisanac, sve je u basni dozvoljeno ali pred kamerama nase nam drage domovine potencijalni robnjari jer nije dozvoljeno svuda i uvijek op vlastitom misljenju govoriti a NAROCIITO PISATI, to ostaje lakse za dokazati

uspuit 23.04.08.05.12

News Portals



- More and more news portals enable user commenting/rating:

News Portals



- More and more news portals enable user commenting/rating:
 - Self-implemented commenting system

News Portals



- More and more news portals enable user commenting/rating:
 - Self-implemented commenting system
 - Social networking site integrated commenting

News Portals



- More and more news portals enable user commenting/rating:
 - Self-implemented commenting system
 - Social networking site integrated commenting
 - Rated comments

News Portals



- More and more news portals enable user commenting/rating:
 - Self-implemented commenting system
 - Social networking site integrated commenting
 - Rated comments
- Great opportunity to study social dynamics!

News Portals



- More and more news portals enable user commenting/rating:
 - Self-implemented commenting system
 - Social networking site integrated commenting
 - Rated comments
- Great opportunity to study social dynamics!
 - *Social systems are systems of communication.* (Niklas Luhman)

Case Study - Croatian News Portals

- Work in progress



Case Study - Croatian News Portals

- Work in progress
- Objective: implementing a sentiment analysis system that will be able to automatically categorize opinions from news portal comments for various purposes (political, cultural, etc. trend analysis)

Case Study - Croatian News Portals

- Work in progress
- Objective: implementing a sentiment analysis system that will be able to automatically categorize opinions from news portal comments for various purposes (political, cultural, etc. trend analysis)
- Done so far:

Case Study - Croatian News Portals

- Work in progress
- Objective: implementing a sentiment analysis system that will be able to automatically categorize opinions from news portal comments for various purposes (political, cultural, etc. trend analysis)
- Done so far:
 - Built generic harvesting application (almost done)

Case Study - Croatian News Portals

- Work in progress
- Objective: implementing a sentiment analysis system that will be able to automatically categorize opinions from news portal comments for various purposes (political, cultural, etc. trend analysis)
- Done so far:
 - Built generic harvesting application (almost done)
 - Harvested over 150000 comments from a number of Croatian news portals

Case Study - Croatian News Portals

- Work in progress
- Objective: implementing a sentiment analysis system that will be able to automatically categorize opinions from news portal comments for various purposes (political, cultural, etc. trend analysis)
- Done so far:
 - Built generic harvesting application (almost done)
 - Harvested over 150000 comments from a number of Croatian news portals
 - 2000 comments chosen at random rated by students for their sentiments (positive/negative, objective/subjective, optimistic/pessimistic)

Case Study - Croatian News Portals

- Work in progress
- Objective: implementing a sentiment analysis system that will be able to automatically categorize opinions from news portal comments for various purposes (political, cultural, etc. trend analysis)
- Done so far:
 - Built generic harvesting application (almost done)
 - Harvested over 150000 comments from a number of Croatian news portals
 - 2000 comments chosen at random rated by students for their sentiments (positive/negative, objective/subjective, optimistic/pessimistic)
 - collected over 72000 ratings

Case Study - Croatian News Portals

- Work in progress
- Objective: implementing a sentiment analysis system that will be able to automatically categorize opinions from news portal comments for various purposes (political, cultural, etc. trend analysis)
- Done so far:
 - Built generic harvesting application (almost done)
 - Harvested over 150000 comments from a number of Croatian news portals
 - 2000 comments chosen at random rated by students for their sentiments (positive/negative, objective/subjective, optimistic/pessimistic)
 - collected over 72000 ratings
 - initial statistical analysis

Case Study - Croatian News Portals

Ocjenjivanje komentara

Dobrodošli Ivek Pivek! Hvala Vam što sudjelujete u istraživanju.
Do sada ste ocjenili o komentara od mogućih 2000.

Komentar: ovog malog netko opasno gura u novine i na sve one "fensi" dogadjaje
--



pohrani

Case Study - Croatian News Portals



- Next steps:

Case Study - Croatian News Portals



- Next steps:
 - Normalize ratings

Case Study - Croatian News Portals



- Next steps:
 - Normalize ratings
 - Vectorize comment text

Case Study - Croatian News Portals

- Next steps:
 - Normalize ratings
 - Vectorize comment text
 - Build support vector machine or other machine learning algorithm to automatically classify comments

Case Study - Croatian News Portals

- Next steps:
 - Normalize ratings
 - Vectorize comment text
 - Build support vector machine or other machine learning algorithm to automatically classify comments
 - Develop mashup site which automatically harvests comments and provides sentiment information / summaries

Case Study - Croatian News Portals

- Next steps:
 - Normalize ratings
 - Vectorize comment text
 - Build support vector machine or other machine learning algorithm to automatically classify comments
 - Develop mashup site which automatically harvests comments and provides sentiment information / summaries
 - Widen the approach to Slovene, Serbian, Bosnian, Macedonian ...

Thank you!



Questions? Comments? Suggestions? Ideas? Critique?

References

1. A. Joshi, A. R. Balamurali and B. Pushpak: Pre-conference workshop on 'Sentiment Analysis' ICON 2009, Hyderabad, 2009.
2. J. Martineau: Identifying and Isolating Text Classification Signals from Domain and Genre Noise for Sentiment Analysis, Ph.D. Diss, UMBC, 2011.
3. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 79–86, 2002.
4. B. Pang and L. Lee: Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, Vol. 2, Nos. 1–2, pp. 1 – 135, 2008.
5. C. Penstein Rosé: Computational Models of Discourse Analysis, available at <http://kanagawa.lti.cs.cmu.edu/11719/sites/default/files/DA-Week9-Lect1.ppt>, 08.05.2012.