



Projektowanie zorientowane na człowieka

Systemy afektywne i analiza sentymenu

Kamil Bortko

Analiza sentymentu

- Komputerowe badanie opinii, sentymentów, ocen, postaw, wpływu, emocji, subiektywności itp. wyrażonych w tekście.
- Przedmiot analiz = materiały tekstowe: recenzje, blogi, dyskusje, aktualności, komentarze, informacje zwrotne, media społecznościowe.
- Alternatywnie określenie to *opinion mining*

Nowe firmy i zastosowania



BUZZLOGIC

LingPipe



brand24
Monitoring Internetu

Platforma Brand24

Brand24

Brand24 to narzędzie do monitoringu Internetu i wsparcia sprzedaży.

To szybki sposób dotarcia do dyskusji związanych z wybraną marką, produktem lub tematem.

Brand24 to także źródło analiz dzięki którym poznasz m.in. najbardziej wpływowych użytkowników social media.



Źródło: Brand24

Platforma Brand24

brand24

Zalogowany jako inerarts@gmail.com | Projekty | Konto | Pomoc | Wyloguj

Projekty

NAZWA NOWE WYNIKI CAŁOŚĆ

nike demo	123 124
adidas	0 101
puma	0 597
apple	69 2908
samsung	79 1576

Pomoc

• pierwsze kroki
• projekt demo
• częste pytania

Korzystaj również z pomocy live dostępnej za pośrednictwem widgetu obecnego poniżej:

Projekt: nike

Panel Podsumowanie

Liczba wyników

Dni Tygodnie Miesiące

Najnowsze wyniki z Social Media

Karol Jarzyna facebook 28 minut temu facebook.com filtruj / usuń

Kasia Rudaś facebook 39 minut temu facebook.com filtruj / usuń

sizeer twitter 59 minut temu twitter.com filtruj / usuń

damonowaa twitter 1 godzinę temu twitter.com filtruj / usuń

Podsumowanie

Wyniki

Analiza

Žródła

Cytaty

infografika ustawienia projektu

1438 wyników ostatnie 7 dni 5% zmiany wyników / tydzień 6936 wyników ostatni miesiąc 41% zmiany wyników / miesiąc

Popularne źródła wyników

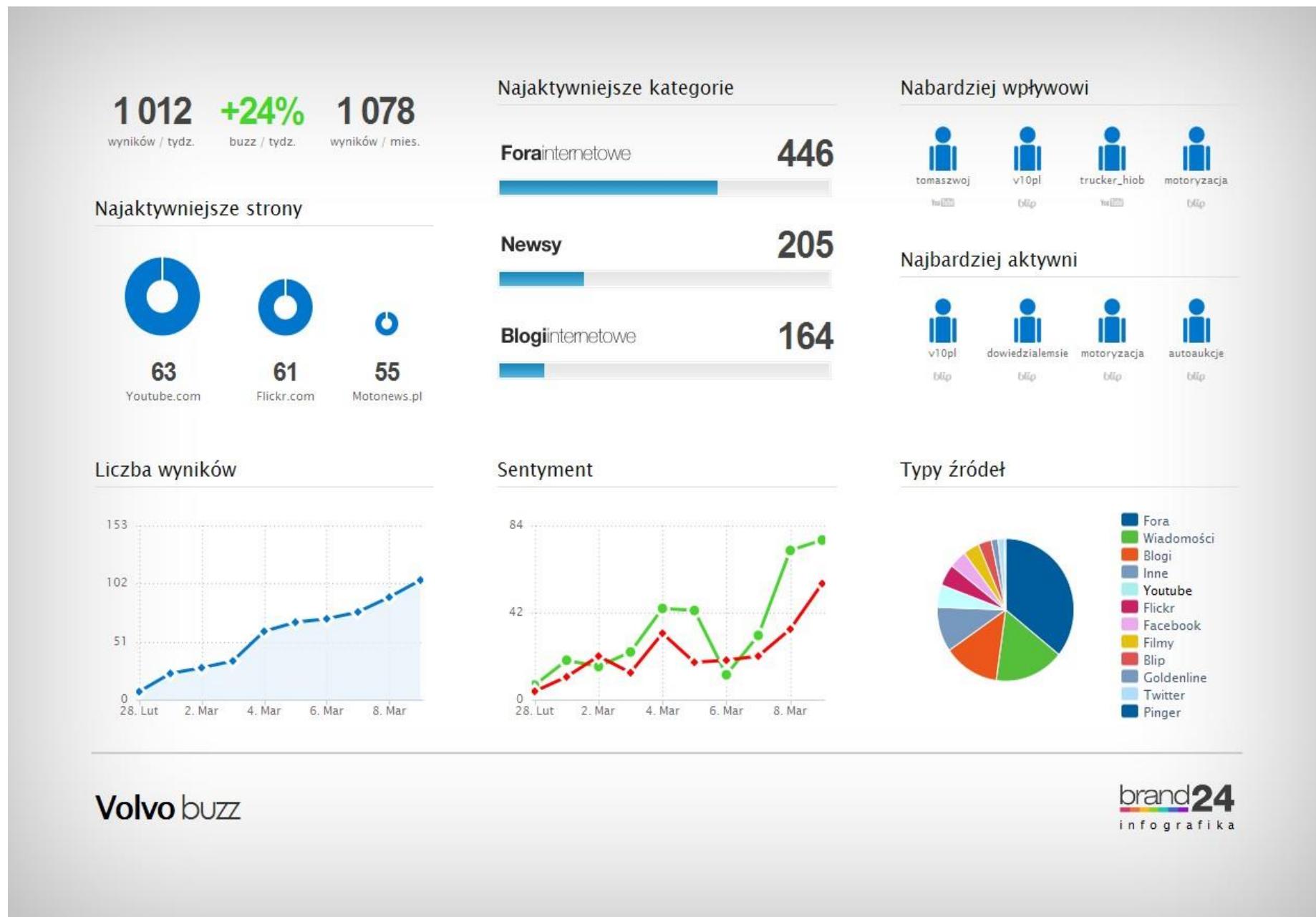
Źródło	Procent
Inne	23.9 %
Twitter	21.7 %
Facebook	20.6 %
Blogi	12.9 %
Forum	8.3 %
Blip	4.3 %
Pinger	2.4 %
Filmy	2.4 %
Wiadomości	1.9 %
Fotosik	1.7 %

Wyniki wg kategorii

Facebook	2823
Blogi	1761
Video	323
Foto	230
Mikroblogi	3879
Forum	1138

Źródło: Brand24

Platforma Brand24



Źródło: Brand24

Platforma Brand24

brand24

Zalogowany jako test5@zule.pl | Monitorowane frazy | Konto | Pomoc | Wyloguj

Frazy

Nazwa	Nowe wyniki
multibank	11
enelmed	93
volvo	22
antyweb	382
samsung	616

Pomoc

- centrum analiz NEW
- fanpage facebook
- pierwsze kroki
- projekt demo
- częste pytania

Korzystaj również z pomocy live dostępnej za pośrednictwem widgetu obecnego poniżej.

Projekt: brand24

Panel > Analiza

Analiza

Podsumowanie Wyniki Źródła Cytaty

infografika ustawienia projektu

Wpływ autorów Social Media

Ostatnie 2 tyg. Cały czas

Autor	Serwis	Wpływ	Opcje
1 Brand24	facebook	326600	info
2 Michał Sadowski	facebook	50776	info
3 hazan	twitter	8336	info
4 Mam Startup	facebook	6612	info
5 Social 360	facebook	5712	info
6 JDomanski	twitter	5436	info
7 hazan	blip	3892	info
8 mediafun	blip	3504	info
9 mediafun	twitter	2906	info
10 Publicon	facebook	2628	info

Popularne źródła wyników

Źródło	Procent
Inne	53.7 %
Facebook	18.1 %
Blogi	12 %
Forum	6.8 %
Twitter	5.6 %
Wiadomości	1.3 %
Filmy	1.2 %
Blip	1.1 %
Pinger	0.1 %

Wyniki wg kategorii

Facebook	238	Blogi	158
Video	18	Foto	0
Mikroblogi	90	Forum	89
Newsy	17	Inne	705

Najaktywniejsi autorzy Social Media

Autor	Serwis	Dotarcie	Wypowiedzi	Opcje
1 Brand24	facebook	1149 osób	142	info
2 Michał Sadowski	facebook	576 osób	44	info
3 brand24pl	twitter	36 osób	28	info

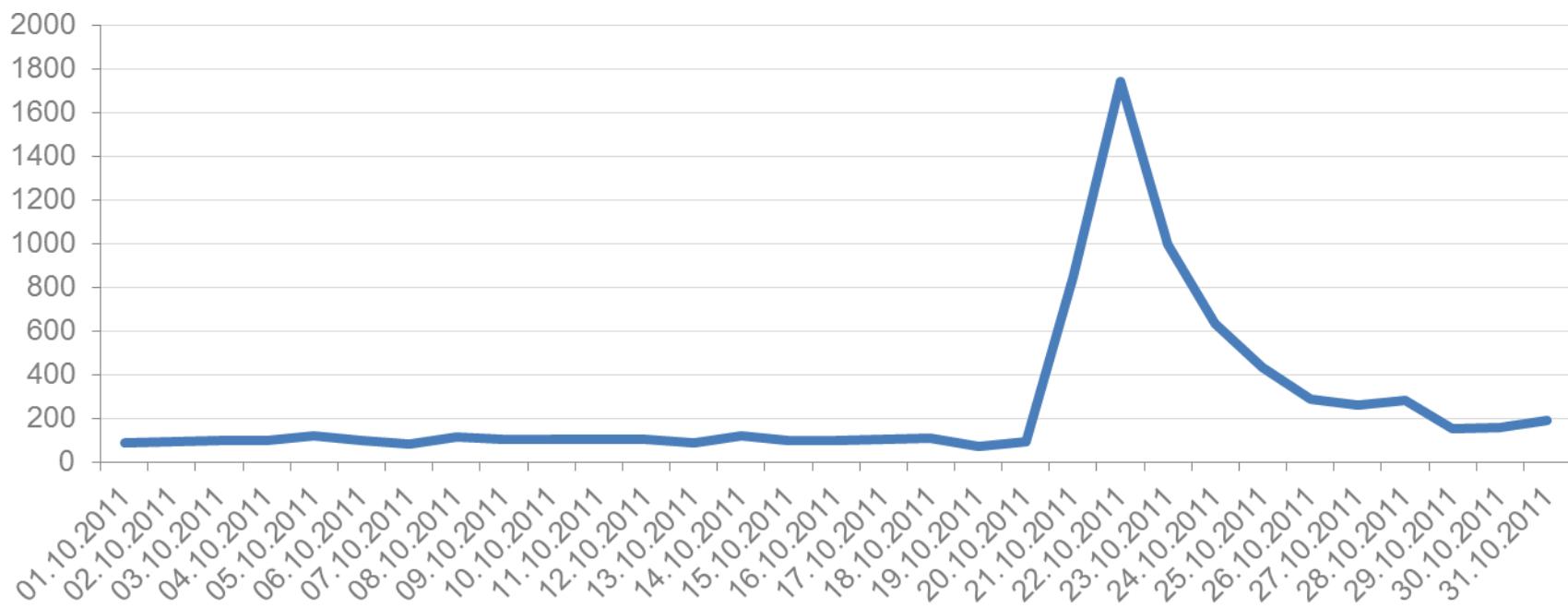
Najaktywniejsze strony

Adres	Wypowiedzi	Opcje
1 facebook.com	281	info
2 mamstartup.pl	226	info
3 twitter.com	75	info

Źródło: Brand24

Platforma Brand24

Liczba treści w podziale na dni – fraza Tesco

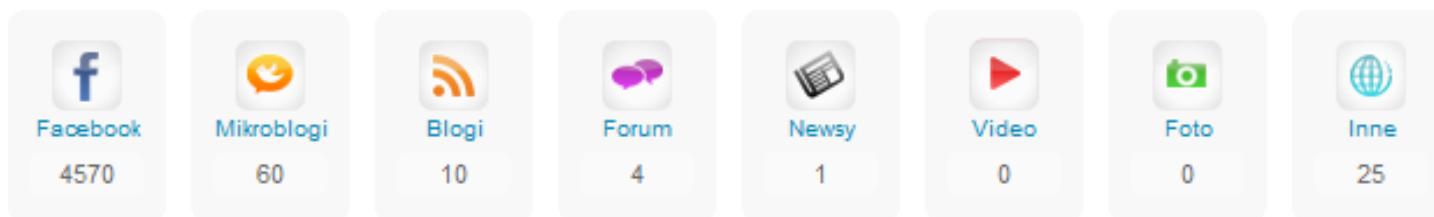


Zamieszczenie filmu wirusowego „Nieemitowana reklama Tesco” spowodowała bardzo dynamiczny przyrost liczby treści dotyczących marki Tesco. O ile w dniach 1-20 października 2011 średnia dzienna liczba treści wynosiła 99, to od 21 października do 31 października średnia dzienna liczba treści wzrosła do 552

Platforma Brand24



Miejsce zamieszczenia treści

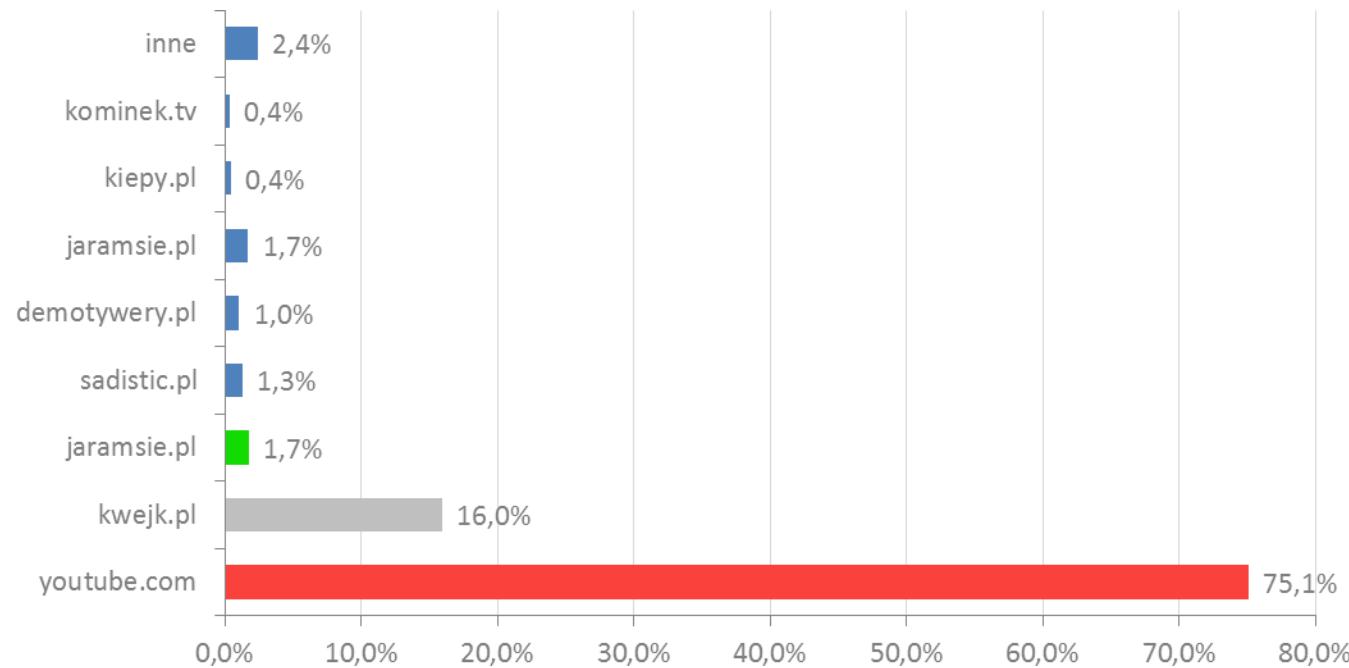


Treści polegały przede wszystkim na zamieszczeniu linku do filmu wirusowego., często opatrzonego komentarzem. 98,4% treści zostało zamieszczonych na Facebooku – wynika to z łatwości rozpowszechniania interesujących treści w tym serwisie, jak i z wielkości tego serwisu (liczby aktywnych internautów)

Źródło: Brand24

Platforma Brand24

Struktura linkowania



Najczęściej treści zawierały linkowanie do dwóch serwisów – YouTube.com oraz kwejk.pl (który faktycznie zawierał zamieszczony widget z YouTube)

Platforma Brand24

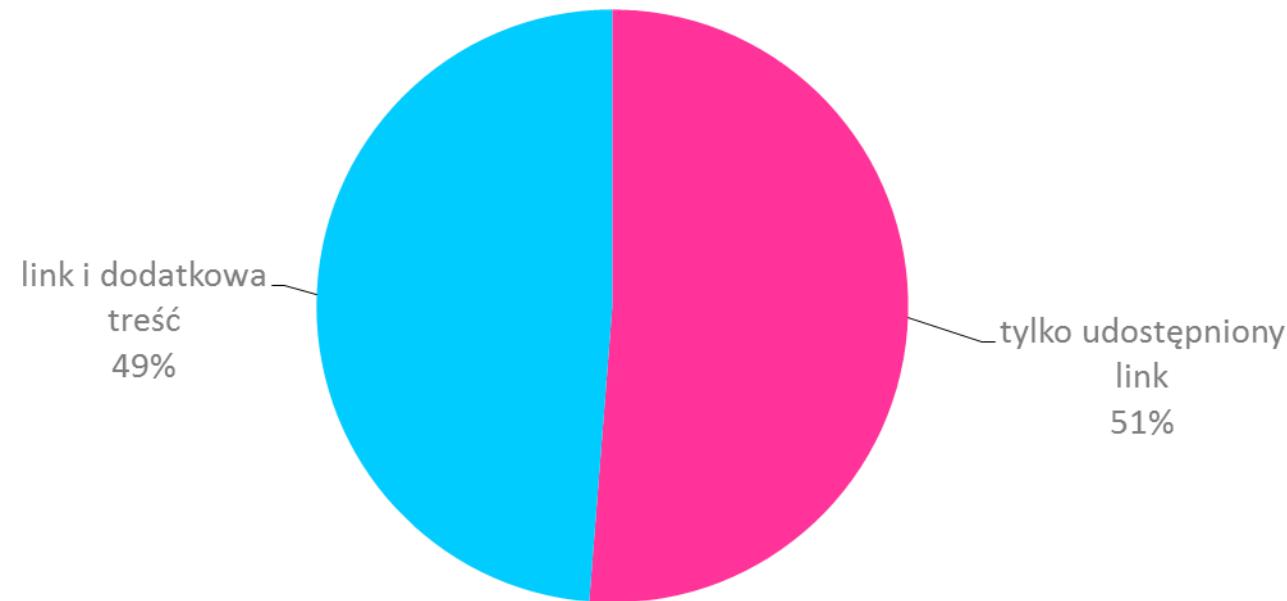
Najbardziej zasięgowi autorzy dla marki Tesco

	Autor	Serwis	Wyników	Dotarcie	Opcje
1	 Tomek Raczek	idź do facebook	1	4950 osób	info
2	 Andrzej Golimont	idź do facebook	1	3920 osób	info
3	 Maciek Florek	idź do facebook	2	3191 osób	info
4	 Caroline✿♥	idź do twitter	1	2276 osób	info
5	 Pani Antywieszakowa	idź do pinger	1	2190 osób	info
6	 Krzysztof Bosak	idź do twitter	1	2153 osób	info
7	 Anna Maria Dębniak	idź do facebook	1	2104 osób	info
8	 Paulina	idź do twitter	1	2063 osób	info
9	 Pixelakaroxi Añ Beats (Tomasz	idź do facebook	1	2063 osób	info
10	 Patrycja (ThePatzae)	idź do twitter	5	2000 osób	info

Źródło: Brand24

Platforma Brand24

Struktura treści

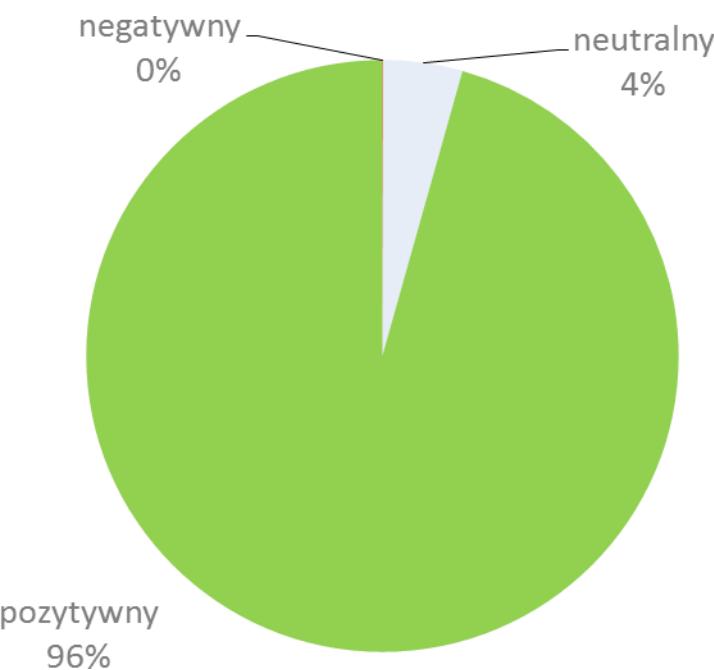


W połowie treści były tylko linkiem do materiału, w połowie zawierały dodatkowe treści wygenerowane przez internautów

Źródło: Brand24

Platforma Brand24

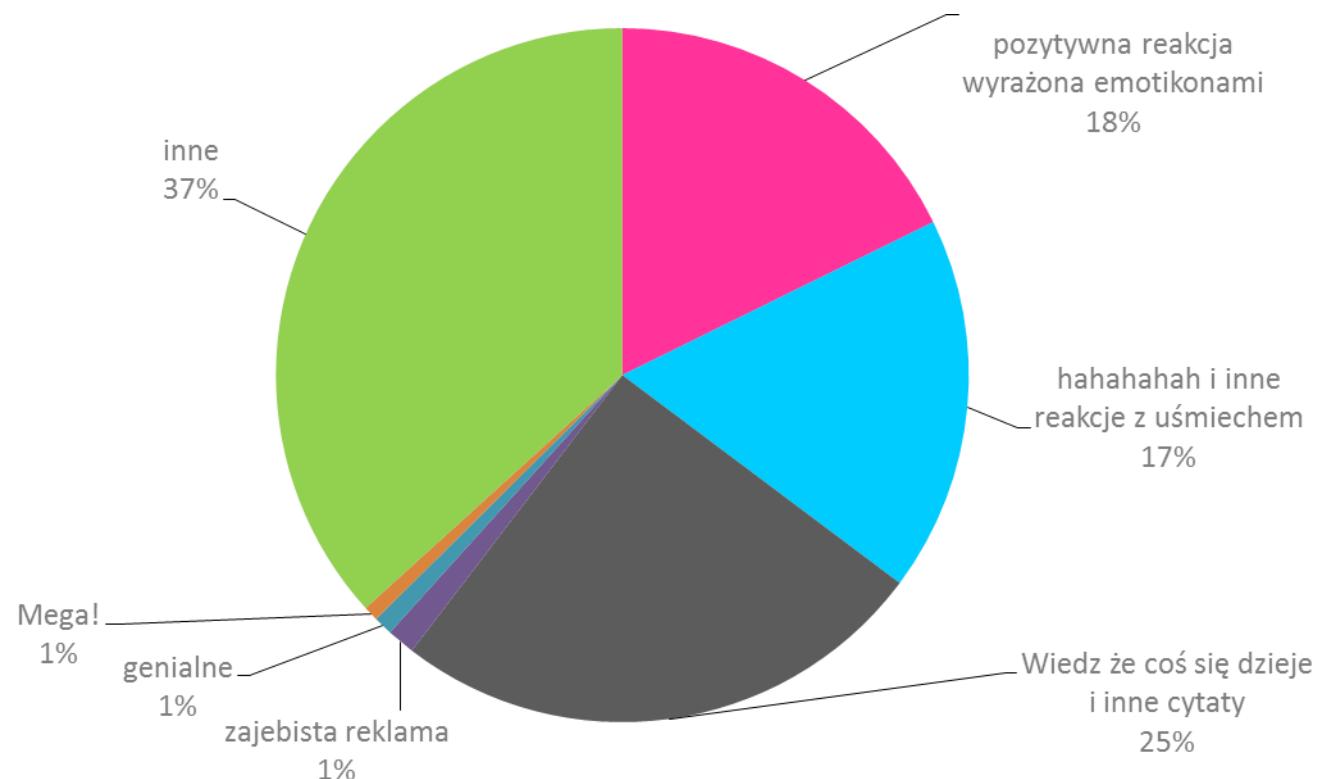
Sentyment treści



Sentyment treści został oceniony dla tych treści, które zawierały wpisy. Wśród tych treści tylko jeden wpis był jednoznacznie negatywny. W tym kontekście należy uznać, że film został oceniony przez internautów bardzo pozytywnie!

Platforma Brand24

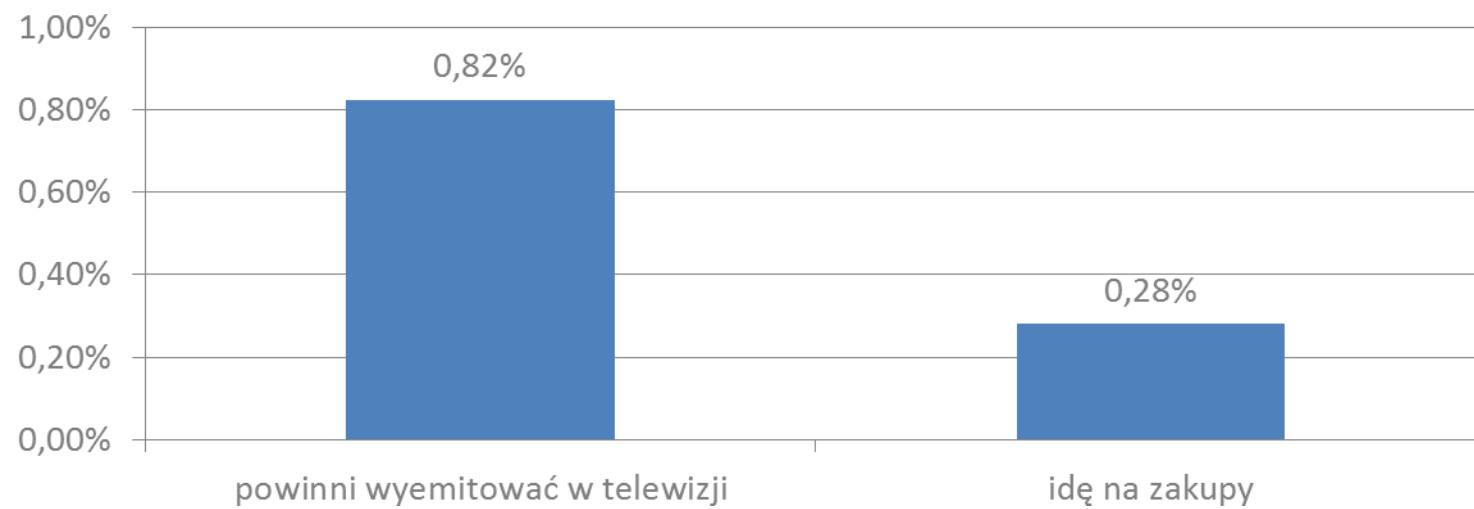
Określenia pozytywne



Źródło: Brand24

Platforma Brand24

Dodatkowe poziomy efektywności



Niektóre wpisy dotyczyły kwestii związanych z emisją reklamy w telewizji, jak i zawierały deklarację chęci zakupów w sklepach Tesco

Platforma Brand24

Źródło: Brand24

Platforma Brand24

CONSUMER INSIGHTS

To chyba w Dove tego nie wiedzą. Mam takie coś (żel pod prysznic) w szarej butelce, ale wielkość otworu jakoś nie współgra z gęstością żelu gi/lub twardością opakowania) - naprawdę trzeba dobrze ścisnąć żeby poleciało. Starcza na dłużej - nie chce mi się tego używać:)



Edyta Peikert ► Frugo Wróć!

11 godz. temu •

Jak mogliście tak zepsuć smak mojego ukochanego zielonego FRUGO!!!! Co robi tam sok grejpfrutowy!!!??? FUJ!!!

Lubię to! • Dodaj komentarz



Radek Marczak no właśnie, ostatnio pilem to takie gorzkie

5 godz. temu przez telefon komórkowy • Lubię to! • 1



Ada Trzop Tak ?

5 godz. temu przez telefon komórkowy • Lubię to!



Edyta Peikert Gdzie nasze mega słodkie zielone frugo????!!!! Jak można było je zepsuć?!

4 godz. temu przez telefon komórkowy • Lubię to!



Napisz komentarz...

Źródło: Brand24

Platforma Brand24

CONSUMER INSIGHTS

Samsung R510 Vista - nie startuje system

Autor	Wiadomość
sitowawa Poziom 1  Dołączył: 11 Cze 2009 Posty: 1 Miasto: Warszawa	D #1 ► 25 Mar 2012 22:16 Samsung R510 Vista - nie startuje system Witam. Mam Samsunga R510 Windows Vista. Po twardym resecie "palcem" nie chce wystartować systemu. Nie idzie w żadnym trybie - ani awaryjnym ani też nie działa Samsung Recovery Solution III. Włącza się niebieski ekran i koniec. Nie idzie z F4, F8 i F12. Poradźcie co robić? Pozdrawiam KS



@SamsungPolska jak wygląda sprawa z sterownikami do Optimus w samsung RF511? Oficjalne nvidii? czy macie swoje na stronie?

12 Sep via TweetDeck ★ Favorite ↗ Retweet ↗ Reply

Źródło: Brand24

Platforma Brand24

OCHRONA REPUTACJI

Marta Rybak
30 sierpnia 2011 o 22:45 · 0

!!!!!!! apeluje do wszystkich konsumentów, aby zwrócić kaszkę bananową „okres przydatności do spożycia do 2012 roku, kod kreskowy: 7613033089 73, ponieważ istnieje możliwość, że zawierają kawałki szkła! Kopuj na status, chociaż nie jesteś rodzicem. Możesz uratować życie jakiegoś dziecka

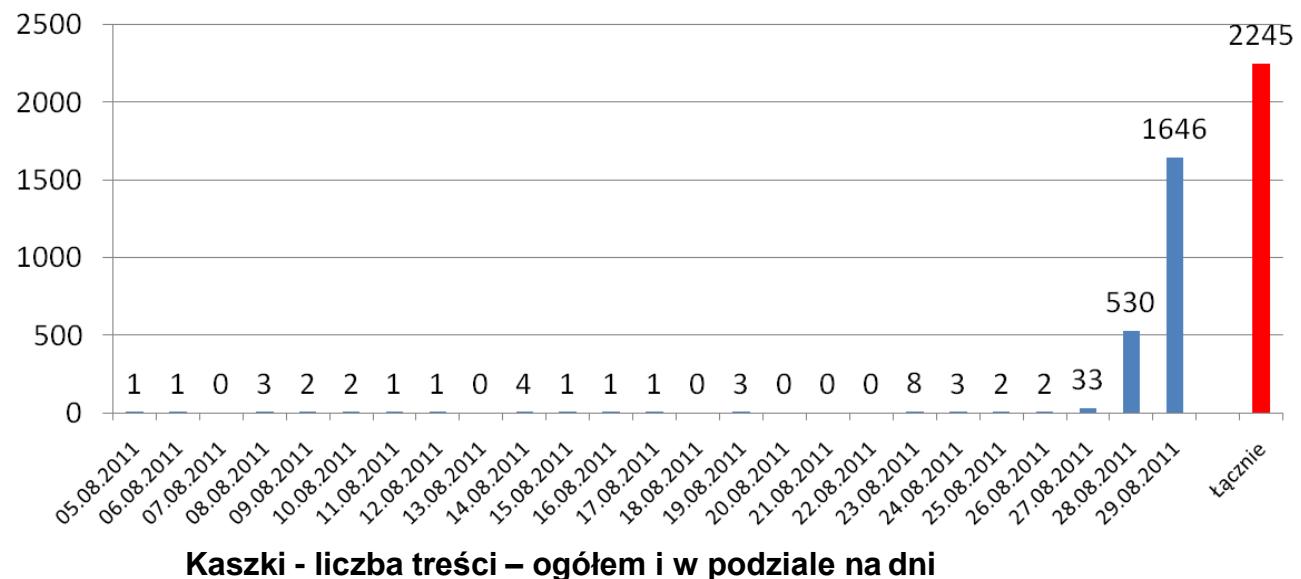
Ostatni wpis: Liczba osób, które go lubią: 2

Agnieszka Dlugosz 27 sierpnia 2011 o 06:58 · 0

Anna Halimak Matko, chłopak sobie jajka, czy że? 1 września 2011 o 08:16 · 0

Marta Rybak m... poważnie !!! Dostępem i przesyłam dalej... 1 września 2011 o 10:59 · 0

Anieta Rybak Jakaż zły nie to ostrzemić nigdy za dużo... 1 września 2011 o 19:43 · 0



Informacja o problemach z kaszkami po raz pierwszy pojawiła się w wypowiedziach Polaków mieszkających w Wielkiej Brytanii oraz Irlandii. Pierwszy odnotowany wpis Polaka dokonany w języku angielskim miał miejsce 5 sierpnia. Pierwszy wpis polskojęzyczny na Facebooku miał miejsce 9 sierpnia i był wpisem Polki mieszkającej w Norwegii.

Właściwa część zaczęła się po godzinie 22:00 26 sierpnia i pierwszym skupiskiem były osoby mieszkające w Poznaniu. Od soboty nastąpił bardzo szybki przyrost wpisów.

Źródło: Brand24

Platforma Brand24

 **Bartosz Kiszewski**
zastanawiam się nad nową taryfą - jaką? i czy dalej orange czy coś innego?
[Udostępnij](#) • 3 września o 20:44 •

 **Marta Lampart** jeżeli chcesz mieć spokój z telefonami od innych osób to przejdź do pley wszyscy będą mieć do ciebie droższe połączenia :)
3 września o 22:06 •

 **Bartosz Kiszewski** zastanawiam się pomiędzy orange a erą
3 września o 22:20

 **Marta Lampart** ery już nie ma :P
3 września o 22:45

 **Marta Lampart** a to co mi wiadomo od znajomych orange ma coraz gorszy zasięg
3 września o 22:46

 **Basia Michalak** orange naj naj :)
4 września o 01:03

 **Bartosz Kiszewski** @Marta: no fakt T-mobile
@Basia: nie wiem czy taki naj :)
4 września o 01:24

 **Bartosz Kiszewski** namawiają mnie na T....
4 września o 11:41

 **Aleksander Roj** Tmobajl to głównie rozmowy w sieci oraz atrakcyjny roaming.
4 września o 12:29

 **Wiola Mączka** T-mobile. Jeszcze chyba najbardziej będzie się opłacał. Sama lada chwila na niego przejdę.
4 września o 14:58

**WSPARCIE
SPRZEDAŻY**



[Profil użytkownika Izabela](#) • Tablica użytkownika Izabela



Izabela Szczepińska

Szukam najlepszej lokaty z oprocentowaniem stałym. Czy ktoś może polecić jakiś dobry Bank?

20 godz. temu

Źródło: Brand24

Social Media Listening Center



Publikacje, metody i algorytmy

Artykuły

Około 17 500 wyników (0,04 s)

Bez ograniczenia czasowego

Od 2020

Od 2019

Od 2016

Zakres niestandardowy...

Wg trafności

Wg daty

Dowolny język

Tylko język polski

uwzględnij patenty

uwzględnij cytaty

Utwórz alert

SemEval-2016 task 4: Sentiment analysis in Twitter

P Nakov, A Ritter, S Rosenthal, F Sebastiani... - arXiv preprint arXiv ..., 2019 - arxiv.org

This paper discusses the fourth year of the "Sentiment Analysis in Twitter Task". SemEval-2016 Task 4 comprises five subtasks, three of which represent a significant departure from previous editions. The first two subtasks are reruns from prior years and ask to predict the ...

☆ 99 Cytowane przez 793 Powiązane artykuły Wszystkie wersje 18 »»

[PDF] arxiv.org

SemEval-2017 task 4: Sentiment analysis in Twitter

S Rosenthal, N Farra, P Nakov - arXiv preprint arXiv:1912.00741, 2019 - arxiv.org

This paper describes the fifth year of the Sentiment Analysis in Twitter task. SemEval-2017 Task 4 continues with a rerun of the subtasks of SemEval-2016 Task 4, which include identifying the overall sentiment of the tweet, sentiment towards a topic with classification on ...

☆ 99 Cytowane przez 598 Powiązane artykuły Wszystkie wersje 17 »»

[PDF] arxiv.org

Semeval-2015 task 10: Sentiment analysis in twitter

S Rosenthal, SM Mohammad, P Nakov, A Ritter... - arXiv preprint arXiv ..., 2019 - arxiv.org

In this paper, we describe the 2015 iteration of the SemEval shared task on Sentiment Analysis in Twitter. This was the most popular sentiment analysis shared task to date with more than 40 teams participating in each of the last three years. This year's shared task ...

☆ 99 Cytowane przez 319 Powiązane artykuły Wszystkie wersje 16 »»

[PDF] arxiv.org

Sentiment analysis in tourism: capitalizing on big data

AR Alaei, S Becken, B Stantic - Journal of Travel Research, 2019 - journals.sagepub.com

Advances in technology have fundamentally changed how information is produced and consumed by all actors involved in tourism. Tourists can now access different sources of information, and they can generate their own content and share their views and ...

☆ 99 Cytowane przez 82 Powiązane artykuły Wszystkie wersje 7

[PDF] researchgate.net

Podobne wyszukiwania

"sentiment analysis" **opinion mining**

"sentiment analysis" **lexicon**

aspect based "sentiment analysis"

"sentiment analysis" **subjectivity**

arabic "sentiment analysis"

naive bayes "sentiment analysis"

deep learning "sentiment analysis"

twitter data "sentiment analysis"

Learning to detect opinion snippet for aspect-based **sentiment analysis**

M Hu, S Zhao, H Guo, R Cheng, Z Su - arXiv preprint arXiv:1909.11297, 2019 - arxiv.org

Aspect-based **sentiment analysis** (ABSA) is to predict the sentiment polarity towards a particular aspect in a sentence. Recently, this task has been widely addressed by the neural attention mechanism, which computes attention weights to softly select words for generating ...

☆ 99 Cytowane przez 2 Powiązane artykuły Wszystkie wersje 3 »»

[PDF] arxiv.org

Computational Intelligence for Affective Computing and **Sentiment Analysis** [Guest Editorial]

E Cambria, S Poria, A Hussain... - IEEE Computational ..., 2019 - ieeexplore.ieee.org

Emotions are intrinsically part of our mental activity and play a key role in communication and decision-making processes. Emotion is a chain of events made up of feedback loops. Feelings and behavior can affect cognition, just as cognition can influence feeling. Emotion ...

☆ 99 Cytowane przez 2 Powiązane artykuły

[PDF] ieee.org

Publikacje, metody i algorytmy

Bez ograniczenia czasowego
Od 2020
Od 2019
Od 2016
Zakres niestandardowy...

Wg trafności
Wg daty

Dowolny język
Tylko język polski

uwzględnij patenty
 uwzględnij cytaty

Utwórz alert

Sentiment analysis of tweets for estimating criticality and security of events

[V Subramanyaswamy](#), [R Logesh](#), [M Abejith](#)... - Improving the Safety ..., 2020 - igi-global.com

Social Media has become one of the major industries in the world. It has been noted that almost three fourth of the world's population use social media. This has instigated many researches towards social media. One such useful application is the sentimental analysis of ...

 99 Cytowane przez 40 Powiązane artykuły Wszystkie wersje 6

[PDF] semanticscholar.org

[HTML] Sentiment Analysis of Conservation Studies Captures Successes of Species Reintroductions

[KS Van Houtan](#), [T Gagne](#), [CN Jenkins](#), [L Joppa](#) - Patterns, 2020 - Elsevier

Learning from the rapidly growing body of scientific articles is constrained by human bandwidth. Existing methods in machine learning have been developed to extract knowledge from human language and may automate this process. Here, we apply sentiment ...

 99 Wszystkie wersje 3

[HTML] sciencedirect.com

Incoherence in Regime Complexes: A Sentiment Analysis of EU-IMF Surveillance

[M Breen](#), [D Hodson](#), [M Moschella](#) - JCMS: Journal of Common ..., 2020 - Wiley Online Library

The proliferation of international institutions means that states can be subject to multiple, overlapping and potentially incoherent international obligations. The regime complexity literature draws attention to this problem but says little about its character and causes. This ...

 99 Powiązane artykuły Wszystkie wersje 3

[PDF] wiley.com

Sentiment Analysis Using Averaged Weighted Word Vector Features

[A Erkan](#), [T Gunor](#) - arXiv preprint arXiv:2002.05606, 2020 - arxiv.org

People use the world wide web heavily to share their experience with entities such as products, services, or travel destinations. Texts that provide online feedback in the form of reviews and comments are essential to make consumer decisions. These comments create ...

 99 Wszystkie wersje 2

[PDF] arxiv.org

Enhancing BERT Representation With Context-Aware Embedding for Aspect-Based Sentiment Analysis

[X Li](#), [X Fu](#), [G Xu](#), [Y Yang](#), [J Wang](#), [L Jin](#), [Q Liu](#)... - IEEE ..., 2020 - ieeexplore.ieee.org

Aspect-based **sentiment analysis**, which aims to predict the sentiment polarities for the given aspects or targets, is a broad-spectrum and challenging research area. Recently, pre-trained models, such as BERT, have been used in aspect-based **sentiment analysis**. This fine ...



[PDF] ieee.org

[HTML] A dynamic group decision making process for high number of alternatives using hesitant Fuzzy Ontologies and sentiment analysis

[JA Morente-Molinera](#), [FJ Cabrerizo](#), [J Mezei](#)... - Knowledge-Based ..., 2020 - Elsevier

The high spread of Internet and social networks have completely changed the way that Group Decision Making methods are designed, developed and implemented. Experts now operate in environments where a large amount of information is available and new ideas ...

 99 Wszystkie wersje 2

[HTML] sciencedirect.com

Implementation of n-gram methodology for rotten tomatoes review dataset sentiment analysis

[P Tiwari](#), [BK Mishra](#), [S Kumar](#), [V Kumar](#) - ..., Methodologies, Tools, and ..., 2020 - igi-global.com

Sentiment Analysis intends to get the basic perspective of the content, which may be anything that holds a subjective supposition, for example, an online audit, Comments on Blog posts, film rating and so forth. These surveys and websites might be characterized into ...

Źródła danych

- **Wewnętrzne dane organizacji**
 - Opinie klientów z wiadomości e-mail, centra obsługi telefonicznej, itp.
- **Aktualności i raporty**
 - Opinie w aktualnościach, komentarzach, raportach, sprawozdaniach firm,
- **Wypowiedzi na stronach WWW**
 - Osobiste doświadczenia i opinie, recenzje, fora, blogi, tweety, mikroblogi, itp
 - Komentarze na temat artykułów, problemów, wydarzeń
 - Komentarze w portalach społecznościowych.

Zastosowania

- **Biznes i organizacje**

- Firmy poszukujące opinii konsumentów korzystały z usług konsultantów, ankiet i grup fokusowych, itp. Teraz mają dane online. Wystarczy je odszukać.
- Kampanie polityczne
- Działania non-profit

- **Osoby fizyczne**

- Podejmowanie decyzji zakupowych
- Opinie o kandydatach politycznych
- Komentarze i opinie dotyczące wydarzeń

- **Marketing**

- Promocja w mediach społecznościowych
- Dobór kanałów
- Reklama targetowana

Poziomy abstrakcji

Id: Abc123 on 5-1-2008 “I bought an *iPhone* a few days ago. It is such a nice *phone*. The *touch screen* is really cool. The *voice quality* is clear too. It is much better than my old *Blackberry*, which was a terrible *phone* and so *difficult to type* with its *tiny keys*. However, *my mother* was mad with me as I did not tell her before I bought the *phone*. She also thought the *phone* was too *expensive*, ...”

- Rozpatrywanie na różnych poziomach:
 - Poziom dokumentu recenzja + lub -?
 - Poziom zdania tzn. czy każde zdanie jest + lub -?
 - Na poziomie cechy produktu usługi

Poziom encji i cech/aspektu

Id: Abc123 on 5-1-2008 “I bought an *iPhone* a few days ago. It is such a nice *phone*. The *touch screen* is really cool. The *voice quality* is clear too. It is much better than my old *Blackberry*, which was a terrible *phone* and so *difficult to type* with its *tiny keys*. However, *my mother* was mad with me as I did not tell her before I bought the *phone*. She also thought the *phone* was too *expensive*, ...”

Typowe elementy:

- Cele/targety opinii: podmioty i ich cechy / aspekty
- Sentymenty: pozytywny i negatywny
- Autorzy opinii: osoby które formułują opinie
- Czas: kiedy opinie są wyrażone

Poziom encji i cech/aspektu

(Jindal i Liu 2006; Liu 2012)

- Regularne opinie : wyrażanie opinii zorientowanej na temat / przedmiot / podmiot:
 - Opinie bezpośrednie:
 - „Ten telefon jest naprawdę super”.
 - Opinie pośrednie:
 - „Po przyjęciu tego leku, poczułem ulgę.”
- Opinie porównawcze:
 - W wypowiedzi więcej niż jedna jednostka. Na przykład „iPhone jest lepszy niż Blacbkerry.”

Struktura opinii

Na opinię składają się podstawowe składniki w postaci uporządkowanej czwórki:

$$(g_i, so_{ijl}, h_i, t_l)$$

gdzie:

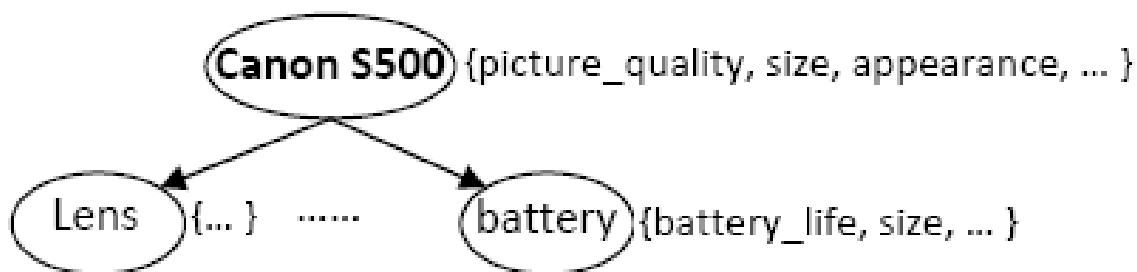
- g_j - obiekt docelowy, przedmiot/cel którego dotyczy opinia
- so_{ijl} - wartość sentymentu wyrażonego przez osobę h_i ukierunkowanego na obiekt docelowy g_j w czasie t_l . so_{ijl} może przyjmować wartości numeryczne reprezentujące sentyment pozytywny, negatywny lub neutralny lub wartość z założonej skali
- h_i - autor opinii
- t_l - czas wyrażenie opinii

Docelowy obiekt opinii

- W niektórych przypadkach, celem opinii jest pojedyncza jednostka lub temat.
 - „*Kocham markę X*” i „*Uwielbiam napój Y*”
- W większości przypadków wypowiedzi są jednak bardziej złożone:
 - “*I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool.*”
 - Celem opinii w zdaniu 3 nie jest ekran dotykowy, ale ekran dotykowy iPhona.
- Jednym z elementów analizy jest dekompozycja obiektu docelowego i wyszczególnienie jego cech

Encje i opinie (Hu i Liu 2004; Liu 2006)

- **Definicja (encji):** Encją jest produkt, osoba, wydarzenie, organizacja, lub temat.
- Encja jest reprezentowana jako hierarchia elementów.
- Każdy węzeł reprezentuje element i jest powiązany ze zbiorem atrybutów.



Opinia może dotyczyć dowolnego węzła lub jego atrybutu. Cechy/aspekty w ramach encji dotyczą i elementów i ich atrybutów

Opinia powiązana z encją (Liu, Ch. 2010)

Opinię można zdefiniować jako uporządkowaną piątkę:

$$(e_i, a_{ij}, o_{ijkl}, h_k, t_l)$$

e_i - określane jest jako encję/byt, którego dotyczy opinia, tj. produkt, usługa, osoba, zdarzenie czy organizacja .

a_{ij} - pewien aspekt j bytu e_i , który może wpływać na opinię innych podmiotów o danym produkcie bądź usłudze, co jest z nim bezpośrednio powiązane.

o_{ijkl} - określa polaryzację opinii sformułowanej w stosunku do aspektu a_{jk} produktu bądź usługi e_i . Może ona przyjąć wartości: pozytywną, negatywną lub neutralną, lub z odpowiednio zdefiniowanymi poziomami intensywności wartości. Przekazuje stopień użyteczności danego produktu czy usługi w określonym aspekcie przez podmiot wyrażający opinię.

h_k - określa podmiot (osobę lub organizację) wyrażający opinię

t_l - czas i moment, w którym dana opinia została wygłoszona

Dekompozycja wpisu

Id: Abc123 on 5-1-2008 “I bought an **iPhone** a few days ago. It is such a nice **phone**. The **touch screen** is really cool. The **voice quality** is clear too. It is much better than my old **Blackberry**, which was a terrible **phone** and so **difficult to type** with its **tiny keys**. However, **my mother** was mad with me as I did not tell her before I bought the **phone**. She also thought the **phone** was too **expensive**, ...”

- Uporządkowana piątka:
 - (iPhone, OGÓLNE, +, Abc123, 5-1-2008)
 - (iPhone, ekran dotykowy, +, Abc123, 5-1-2008)

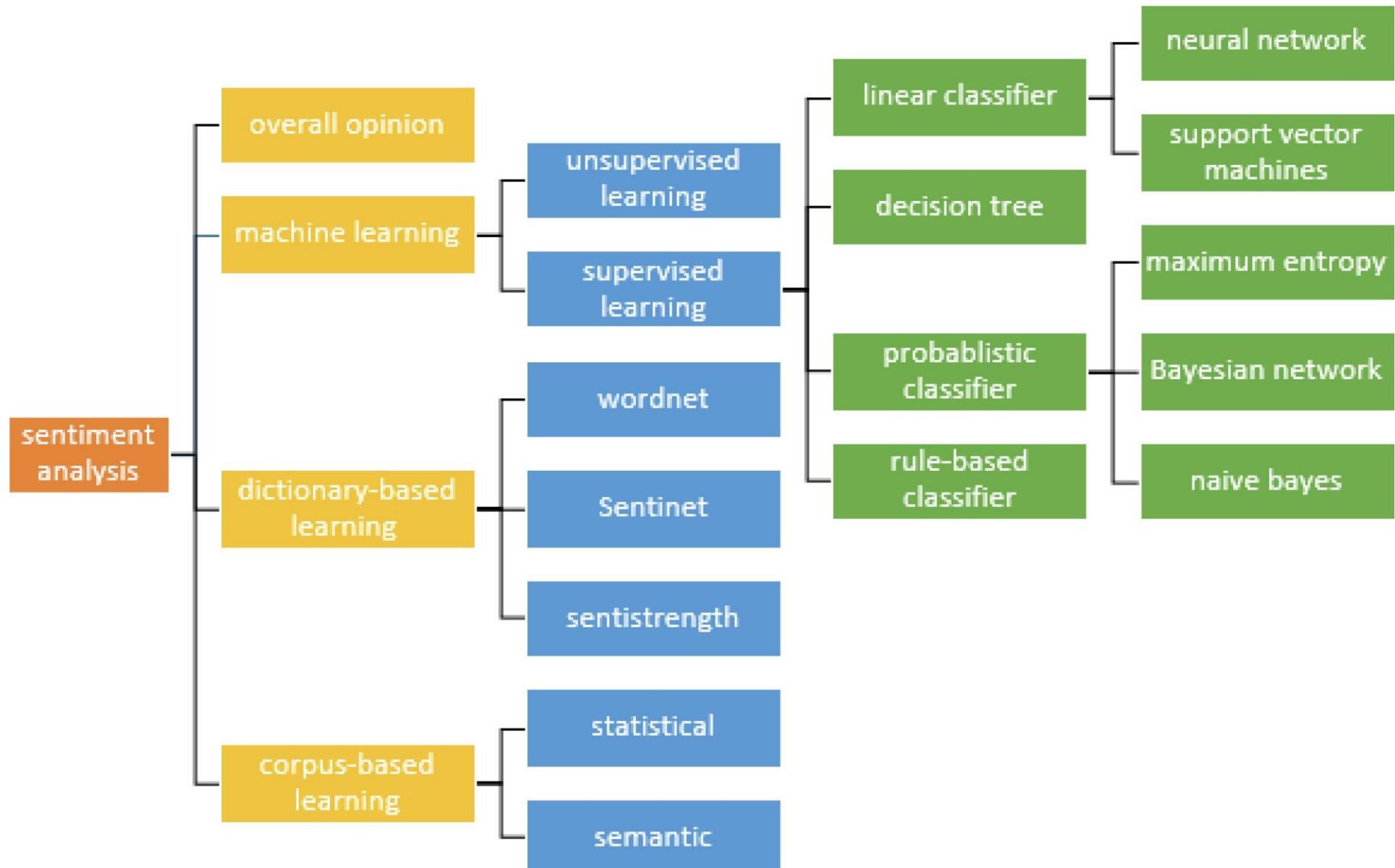
Subiektywizm vs obiektywizm

- Wypowiedź obiektywna przedstawia rzeczowe informacje.
- Zdanie subiektywne wyrażą osobiste opinie, odczucia uczucia, lub emocje.
- Subiektywne wyrażenia mają różne formy, np. opinie, zarzuty, pragnienia, przekonania, podejrzenia, i spekulacje (Wiebe, 2000; Riloff i in 2005).
- Subiektywna wypowiedź może zawierać pozytywne lub negatywne opinie.
- Większość opinii ma elementy subiektywne powiązane z faktami (Liu, 2010)

Oceny racjonalne i emocjonalne

- **Ocena racjonalna**
- Oceny racjonalne nie zawierają przekazu emocjonalnego
 - np. „Wyświetlacz jest jasny”
- **Ocena emocjonalna**
 - „Kocham ten telefon”
 - „Głos jest kryształowo czysty”

Kierunki rozwoju metod



Chakriswaran, P., Vincent, D. R., Srinivasan, K., Sharma, V., Chang, C. Y., & Reina, D. G. (2019). Emotion AI-Driven Sentiment Analysis: A Survey, Future Research Directions, and Open Issues. *Applied Sciences*, 9(24), 5462.

Metody leksykonowe

- Zbiory słów opiniujących lub całe korpusy tekstów, występowanie słów i wyrażenia kluczowe.
- Analizie poddaje się zawartość tekstów i wypowiedzi, wybierając istotne słowa i frazy.
- Zakłada się, że istnieją pewne słowa, które ludzie często używają do wyrażania silnych uczuć. Wystarczy zatem stworzyć listę takich słów i polegać wyłącznie na nich, aby sklasyfikować dany tekst [LEE, 2002].
- Podejście wykorzystujące metody słownikowe wykorzystuje znaczenia analizowanych wyrazów i bierze pod uwagę reguły leksykalne danego języka.

Metody leksykonowe

- Duża zależność od języka i specyfiki zastosowań
- Niezbędna jest znajomość gramatyki danego języka oraz specyfiki wypowiedzi związanej ze stosowanym słownictwem
- Słowniki dla różnych języków
- Podatność metod na różne interpretacje słownikowe

Metody leksykonowe

[PDF] A lexicon based sentiment analysis retrieval system for tourism domain

[PDF] agrilife.org

A García, S Gaines, MT Linaza - Expert Syst Appl Int J, 2012 - agrilife.org

Sentiment analysis has been extensively investigated during the last years mainly for English language. Currently, existing approaches can be split into two main groups: methods **based** on the combination of **lexical** resources and Natural Language Processing ...

☆ 99 Cytowane przez 30 Powiązane artykuły Wszystkie wersje 7 »»

Lexicon-based sentiment analysis of teachers' evaluation

[PDF] hindawi.com

Q Rajput, S Haider, S Ghani - Applied Computational Intelligence and ..., 2016 - hindawi.com
The end of the course evaluation has become an integral part of education management in almost every academic institution. The existing automated evaluation method primarily employs the Likert scale **based** quantitative scores provided by students about the delivery ...

☆ 99 Cytowane przez 30 Powiązane artykuły Wszystkie wersje 7 »»

Lexicon-based sentiment analysis for reviews of products in Brazilian Portuguese

LV Avanço, MGV Nunes - 2014 Brazilian Conference on ..., 2014 - ieeexplore.ieee.org

This paper presents some results on **lexicon-based** classification of **sentiment** polarity in web reviews of products written in Brazilian Portuguese. They represent a first step towards a robust opinion miner from reviews of technology products. The evaluation shows the ...

☆ 99 Cytowane przez 32 Powiązane artykuły Wszystkie wersje 6

A generation model to unify topic relevance and **lexicon-based sentiment** for opinion retrieval

[PDF] acm.org

M Zhang, X Ye - Proceedings of the 31st annual international ACM ..., 2008 - dl.acm.org

... **based** ranking criterion now serves as the weighting factor for the **lexicon-based sentiment** ranking function ... In addition, the **lexicons** used in this work are all domain-independent ones ... Dictionaries are constructed according to existing **lexical** categories [6, 10, 19] or the word ...

☆ 99 Cytowane przez 130 Powiązane artykuły Wszystkie wersje 11

Lexicon-based sentiment analysis: Comparative evaluation of six **sentiment lexicons**

[PDF] ntu.edu.sg

CSG Khoo, SB Johnkhan - Journal of Information Science, 2018 - journals.sagepub.com

This article introduces a new general-purpose sentiment lexicon called WKWSCI Sentiment Lexicon and compares it with five existing lexicons: Hu & Liu Opinion Lexicon, Multi-perspective Question Answering (MPQA) Subjectivity Lexicon, General Inquirer, National ...

☆ 99 Cytowane przez 36 Powiązane artykuły Wszystkie wersje 7

Lexicon-based sentiment analysis of Facebook comments in Vietnamese language

S Trinh, L Nguyen, M Vo, P Do - Recent developments in intelligent ..., 2016 - Springer

Social media websites like Twitter, Facebook etc. are a major hub for users to express their opinions online. **Sentiment analysis** which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in blog posts, comments ...

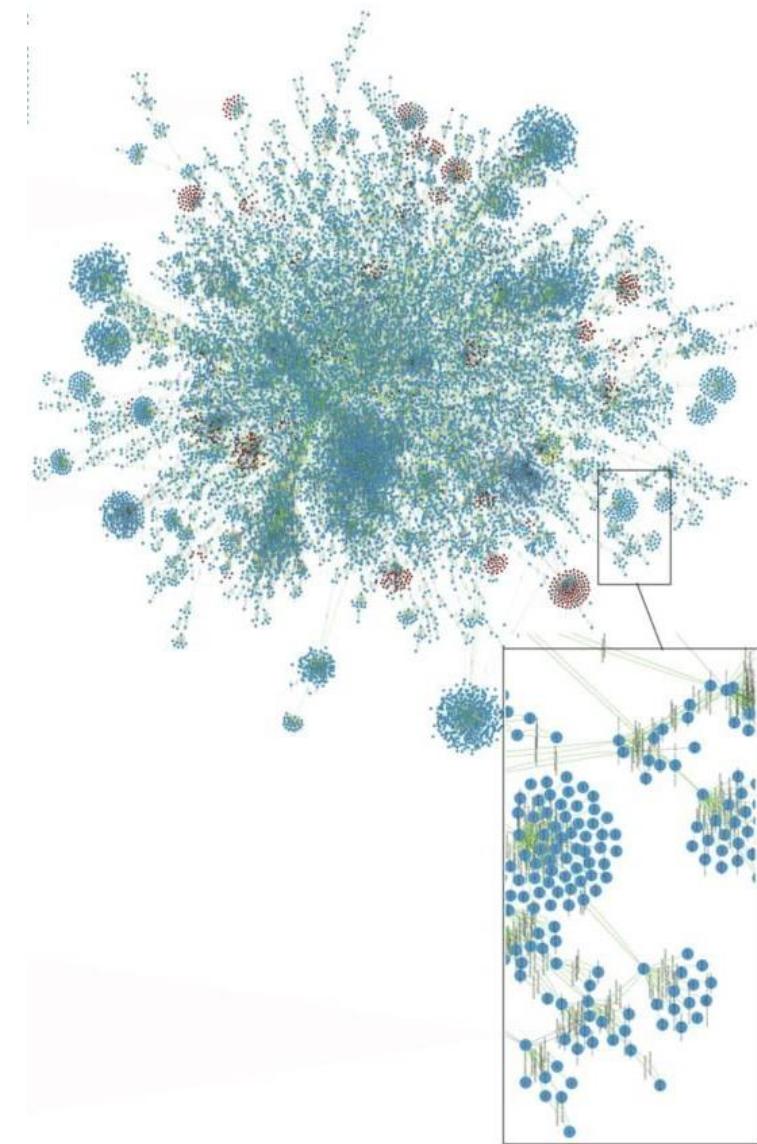
☆ 99 Cytowane przez 22 Powiązane artykuły Wszystkie wersje 3

Metody leksykonowe

- Słowa lub zwroty powiązane z wydźwiękiem (polaryzacja)
 - Pozytywny: piękny, wspaniały, dobry, niesamowity,
 - Negatywny: zły, straszny, brudny.
- Zależne od kontekstu, a nie tylko od dziedziny
- Główne sposoby budowania list :
 - Na podstawie słowników
 - Oparte na korpusach
 - Manualne

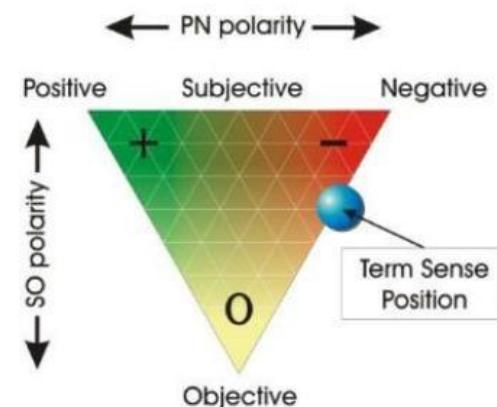
WordNet

- WordNet - elektroniczna baza leksykalna, jest uważana za najważniejszy zasób dostępny dla naukowców w dziedzinie lingwistyki obliczeniowej, analizy tekstu i wielu powiązanych dziedzin.
- Konstrukcja jest inspirowana obecnymi psycholingwistycznymi i obliczeniowymi teoriami ludzkiej pamięci leksykalnej.
- Angielskie rzeczowniki, czasowniki, przymiotniki i przysłówki są zorganizowane w zestawy synonimów, z których każdy reprezentuje jedną leżącą u podstaw leksykalizowaną koncepcję.
- Zestawy synonimów łącz różne relacje w postaci sieci.



SENTIWORDNET

- SentiWordNet 3.0, zasób leksykalny opracowany specjalnie do obsługi klasyfikacji opinii i eksploracji opinii.
- SentiWordNet 3.0 jest publicznie dostępny do celów badawczych, obecnie licencjonowany dla ponad 300 grup badawczych i wykorzystywany w różnych projektach badawczych na całym świecie.
- SentiWordNet jest wynikiem automatycznego dodawania adnotacji do wszystkich zestawów WORDNET zgodnie z ich stopniami pozytywności, negatywności i neutralności.
- Algorytm został użyty do klasyfikacji obiektów WORDNET
- Poprawa dokładności o około 20% w stosunku do SentiWordNet 1.0.



Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In Lrec (Vol. 10, No. 2010, pp. 2200-2204).

SENTIWORDNET



SłowoSieć

PLWORDNET SŁOWOSIEĆ

wielka sieć wyrazów

191 000 słów
285 000 znaczeń
ponad 600 000 relacji
255 000 haseł polsko-angielskich
80 000 jednostek z anotacją emocjonalną
darmowa licencja
największy wordnet na świecie

Projekty, w ramach których była lub jest rozwijana SłowoSieć:

Automatyczne metody konstrukcji sieci semantycznej leksemów polskich na potrzeby przetwarzania języka naturalnego (2005–2008),

Konstrukcja zasobów leksykalnych przez rozpoznawanie relacji semantycznych na podstawie danych morfosyntaktycznych i semantycznych w korpusach tekstu (2009–2012), finansowany przez

NEKST — Adaptacyjny system wspomagający rozwiązywanie problemów w oparciu o analizę treści dostępnych źródeł elektronicznych (2010–2013),

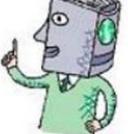
SyNaT — Zadanie badawcze „Utworzenie uniwersalnej, otwartej, repozytoryjnej platformy hostingowej i komunikacyjnej dla sieciowych zasobów wiedzy dla nauki, edukacji i otwartego społeczeństwa wiedzy” (2010–2013)

CLARIN-PL - Polska część infrastruktury naukowej CLARIN ERIC: Wspólne zasoby językowe i infrastruktura technologiczna (2013–2015)

Źródło: <http://plwordnet.pwr.wroc.pl/wordnet/>

Typowe zastosowania

Pozyskiwanie z danych tekstowych nastawienia odbiorców, klientów, opinii, preferencji politycznych.



OPINE
Ana-Maria Popescu, Bao Nguyen, Oren Etzioni

Home | Language: English ▾
[New York City hotels](#) > Renaissance New York Hotel Times Square

Review Summary

Staff: [excellent \(7\)](#), [great \(3\)](#), [very helpful \(2\)](#), [poor](#), [fantastic](#), [helpful](#), [love](#), [good](#), [view all \(17\)](#)

Location: [great \(4\)](#), [best \(3\)](#), [good \(2\)](#), [fabulous](#), [fantastic](#), [ideal](#), [superb](#), [not great](#), [love](#), [view all \(15\)](#)

Room: [nice \(5\)](#), [great \(2\)](#), [not great \(2\)](#), [good \(2\)](#), [very nice \(2\)](#), [excellent](#), [superb](#), [lovely](#), [average](#), [view all \(17\)](#)

Quality: [best](#), [fantastic](#), [lovely](#), [recommend](#), [love](#), [nice](#), [fine](#), [view all \(7\)](#)

Food: [very good \(2\)](#), [fantastic](#), [lovely](#), [not great](#), [great](#), [view all \(6\)](#)

Bathroom beauty: [beautiful](#)

Bar: [fabulous](#), [great](#), [view all \(2\)](#)

Staff friendliness: [friendly \(4\)](#), [very friendly \(2\)](#), [incredibly friendly](#), [unfriendly](#), [view all \(8\)](#)

Room bed comfort: [comfy \(2\)](#), [comfortable \(2\)](#), [extremely comfortable](#), [view all \(5\)](#)

Bathroom: [great \(2\)](#), [elegant](#), [very nice](#), [nice](#), [view all \(5\)](#)

Room cleanliness: [clean \(2\)](#)

User comments:

the rooms were clean and smelled great . [Read more](#)

The rooms were clean, spacious, soundproof and well-appointed . [Read more](#)

Metody oparte na korpusie

- Polegają na wzorcach składniowych większych fragmentów tekstu.
- Metody: Hazivassiloglou and McKeown, 1997; Turney, 2002; Yu and Hazivassiloglou, 2003; Kanayama and Nasukawa, 2006; Ding, Liu and Yu, 2008
- Reprezentacja zależna od dziedziny
- Przydzielanie orientacji / polaryzacji do słowa / frazy

Algorytm Turneya

Uczenie częściowo nadzorowane, oparte na tworzeniu słowników całych wyrażeń. W eksperymencie jako dane wejściowe zostały wykorzystane wyrazy pozytywne i negatywne oraz lista sekwencji części mowy, będącymi wzorcami do dopasowywania potencjalnych wyrażeń posiadających wydźwięk.

Badania były przeprowadzone na recenzjach z portalu Epinions. Eksperymenty takie nad wydźwiękiem całych wyrażeń są istotne, ze względu na wieloznaczność wyrazów – często bez kontekstu nie jest możliwe określenie wydźwięku słowa

Pierwszym krokiem algorytmu jest oznaczenie części mowy wyrazów w tekście oraz wybór potencjalnie interesujących wyrażeń zgodnie z listą sekwencji części mowy.

Następnie wykonywane są obliczenia w celu określenia współczynnika orientacji semantycznej (ang. semantic orientation), który oparty jest na współwystępowaniu wyrażenia z wyrazami pozytywnymi i negatywnymi.

Turney, P. D., & Littman, M. L. (2002). Unsupervised learning of semantic orientation from a hundred-billion-word corpus. arXiv preprint cs/0212012.

Algorytm Turneya

Współczynnik orientacji semantycznej korzysta z miary informacji wzajemnej, która określa prawdopodobieństwo współwystępowania dwóch słów:

$$PMI(slowo_1, slowo_2) = \log_2 \frac{p(slowo_1 \& slowo_2)}{p(slowo_1)p(slowo_2)}$$

- W przypadku algorytmu Turneya obliczono informację wzajemną współwystępowania wyrażenia w pobliżu słowa pozytywnego oraz współwystępowania wyrażenia ze słowem negatywnym.
- Współczynnik orientacji semantycznej jest to różnica tych informacji wzajemnych.
- Po jego obliczeniu można ocenić, że wyrażenia o współczynniku większym od zera mają wydźwięk pozytywny a mniejszym od zera negatywny.
- Jakość tego algorytmu w dużej mierze zależy od dobrania parametrów, definicji współwystępowania wyrażenia z wyrazem oraz doboru wyrazów pozytywnych i negatywnych do przetwarzanych tekstów.
- Poprzez dopasowywanie do tekstu całych wyrażeń rośnie złożoność późniejszego wykorzystania takiego słownika.

Turney, P. D., & Littman, M. L. (2002). Unsupervised learning of semantic orientation from a hundred-billion-word corpus. arXiv preprint cs/0212012.

Algorytm Turneya

SO-PMI-IR (Semantic Orientation from Pointwise Mutual Information and Information Retrieval)

$$\text{SO-PMI-IR}(word) = \text{PMI}(word, \{\text{positive paradigms}\}) - \text{PMI}(word, \{\text{negative paradigms}\})$$

{good, nice, excellent, positive, fortunate, correct, and superior}
{bad, nasty, poor, negative, unfortunate, wrong, and inferior}

- Inne eksperymenty wykorzystują wyszukiwarkę AltaVista i zaindeksowane około 350 milionów stron internetowych w języku angielskim. Biorąc pod uwagę szacunek 300 słów na stronę internetową, to reprezentuje korpus zawierający co najmniej sto miliardów słów.
- W tabeli wykorzystano siłę orientacji semantycznej jako miarę pewności, że słowo zostanie poprawnie sklasyfikowane. Testowe słowa zostały posortowane w porządku malejącym wartość ich orientacji semantycznej i słowa o najwyższym rankingu były sklasyfikowane.

Percent of full test set	Size of test set	Accuracy
100%	3596	79.70%
75%	2697	86.43%
50%	1798	90.04%
25%	899	92.21%

Metody statystyczne

- Tekst traktowany jest obiektem.
- Charakteryzuje się go za pomocą danych ilościowych, które odzwierciedlają zawartość dokumentu np. liczbę słów lub fraz.
- Tekst–obiekt reprezentowany jest w postaci wektora w wielowymiarowej przestrzeni wyznaczonej przez zbiór cech (ang. *features*) opisujących dokumenty.
- Trzy etapy : tokenizacja (podział) tekstu, ekstrakcja cech na podstawie których zostanie wytrenowany klasyfikator oraz etapu klasyfikacji.

Metody statystyczne

- Algorytmy z dziedziny uczenia maszynowego takie jak klasyfikator Bayesowski, maszyna wektorów nośnych SVM czy sieci neuronowe.
- Mogą one uwzględniać współwystępowanie słów, które może być przeprowadzone poprzez reprezentację dokumentu jako nieuporządkowanej liczby słów.
- Algorytmy te nie wymagają słowników ani definicji gramatyki, operują jedynie na danych trenujących (i ewentualnie testowych do ewaluacji uczenia)
- Osiągają one słabe wyniki w przypadku krótkich tekstów, pojedynczych zdań czy fraz.

Klasyfikacja dokumentów

bing

HP printer

ALL RESULTS

Shopping

POPULAR FEATURES

- all
- Affordability
- Speed
- Print Quality
- Reliability
- Ease Of Use
- Brand
- Installation
- Size
- Compatibility

SHOPPING

HP LaserJet 1020 - printer - B/W - laser, 15ppm, USB

from \$179 (2 stores)  Bing cashback · 3%



★★★★☆ user reviews (177)

The HP LaserJet 1020 Printer, an excellent laser printer for the cost-conscious consumer. It offers high-quality LaserJet printing in a compact size, and at a price you can afford.

[user reviews](#) [product details](#) [expert reviews](#) [compare prices](#)

user reviews view: **positive comments (44)**

speed  96%

The quality is as good as any laserjet printer I've used and the speed is fast.
Love Reading www.amazon.com 3/17/2006 more...

Quick and fast transaction.
Arthur L. Taylor www.amazon.com 2/5/2008 more...

It's small and fast and very reliable.
Muffinhead's mom www.amazon.com 1/9/2007 more...

Opinie aspektowe

Google products

Sony Cyber-shot DSC-W370 14.1 MP Digital Camera (Silver)

[Overview](#) - [Online stores](#) - [Nearby stores](#) - [Reviews](#) - [Technical specifications](#) - [Similar items](#) - [Accessories](#)

 \$140 [online](#), \$170 [nearby](#)
 159 reviews  0

Reviews

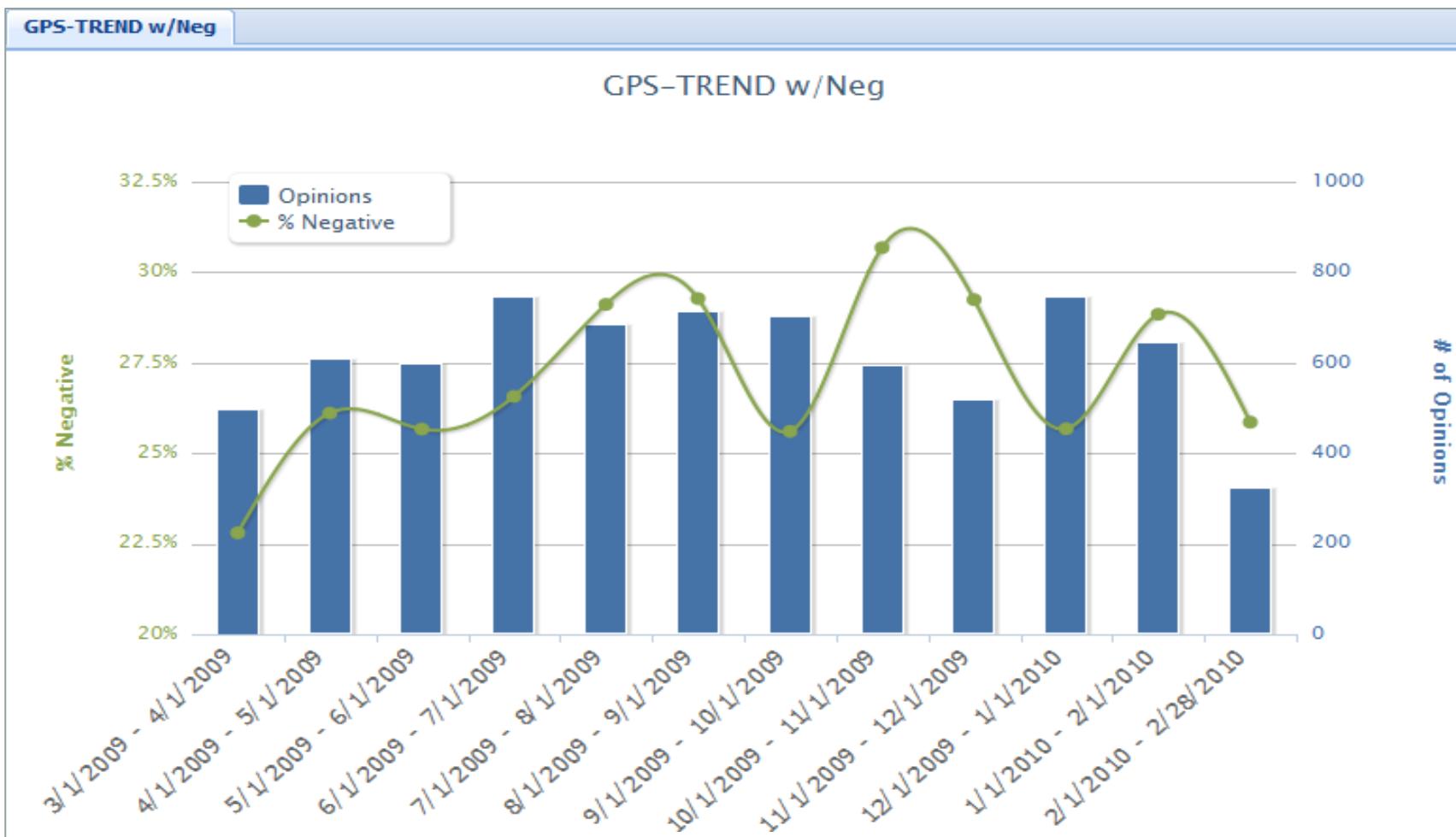
Summary - Based on 159 reviews

What people are saying

pictures	 	"We use the product to take quickly photos."
features	 	"Impressive panoramic feature."
zoom/lens	 	"It also record better and focus better on sunny days."
design	 	"It has the slightest grip but it's sufficient."
video	 	"Video zoom is choppy."
battery life	 	"Even better, the battery lasts long."
screen	 	"I Love the Sony's 3" screen which I really wanted."

Opinie aspektowe i agregacja



Amazon reviews

248 of 263 people found the following review helpful:

★★★★★ **This is one to get if you want 5MP**, April 14, 2004

By [Gadgester "No Time, No Money"](#) (Mother Earth) - [See all my reviews](#)

TOP 100 REVIEWER

[Amazon Verified Purchase](#) ([What's this?](#))

This review is from: [Canon PowerShot S500 5MP Digital Elph with 3x Optical Zoom \(Electronics\)](#)

The new Canon PowerShot S500 is a 5MP upgrade to the immensely popular S400 model, which was a 4MP digital camera. The S500 produces excellent images, is easy to use, and is compact enough to carry in a pocket. 3X optical zoom is standard on these cameras. Besides shooting still photos, you can record low-res video clips as well as audio clips, but don't expect high quality on either.

For a hundred bux less, you can get the 4MP S410 model which is otherwise identical to the S500. Should you go for this or the S410? I think for most consumers 4MP is plenty enough, with room for cropping and enlargements. 5MP is only necessary if you really crop a lot *and* plan to blow up the cropped images. The S410 strikes a great balance between pixel count and price -- it's a better value.

Help other customers find the most helpful reviews

[Report abuse](#) | [Permalink](#)

Was this review helpful to you?

[Comment](#)

41 of 41 people found the following review helpful:

★★★★☆ **E18 Error / problem with the lens**, September 29, 2004

By [Johnathan Parker](#) (Springdale, AR USA) - [See all my reviews](#)

REAL NAME

This review is from: [Canon PowerShot S500 5MP Digital Elph with 3x Optical Zoom \(Electronics\)](#)

This is my second Canon digital elph camera. Both were great cameras. Recently upgraded to the S500. About 6 months later I get the dreaded E18 error. I searched the Internet and found numerous people having problems. When I determined the problem to be the lens not fully extending I decided to give it a tug. It clicked and the camera came on,

Uczenie nadzorowane

- **Klasyfikacja tekstu**
- **Zbiory treningowe**
 - Recenzje filmów z ocenami, opinions, amazom
 - Ocena 4–5 pozytywna
 - Ocena 1-2 negatywny
- Neutralny wydźwięk zazwyczaj zignorowany
- SVM daje najlepszą dokładność klasyfikacji na podstawie na zrównoważony trening dane
 - Typowy wynik: 80–90% precyzja

Budowanie zbiorów uczących

- Dla pozyskanego zbioru komentarzy losowanie próby
- Udział respondentów do manualnej klasyfikacji (ang. annotations)
- Przypisywanie wartości (neutralna, pozytywna, negatywna).
- Analiza stabilności dla respondentów
- Iteracje z powtórzeniami
- Wykorzystanie zbioru uczącego do klasyfikacji

Procedura weryfikacji metod

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives},$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives},$$

$$F-measure = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)}.$$

Precision - procent poprawnie określonego wydżwięku przez metodę w stosunku do sumy wszystkich wartości przypisanych do danej polarności.

Recall - wskazuje stosunek liczby opinii o poprawnie określonym wydżwięku do całkowitej liczby wszystkich komentarzy, które rzeczywiście mają taką wartość wydżwięku.

F-measure - określa efektywność analizy wydżwięku w oparciu o parametry Recall i Precision. F-measure jest to de facto średnia ważona obu tych parametrów.

True Positives - liczebność prawdziwych przypisań pozytywnego wydżwięku, czyli takich, w których metoda i respondenci byli zgodni w ocenie.

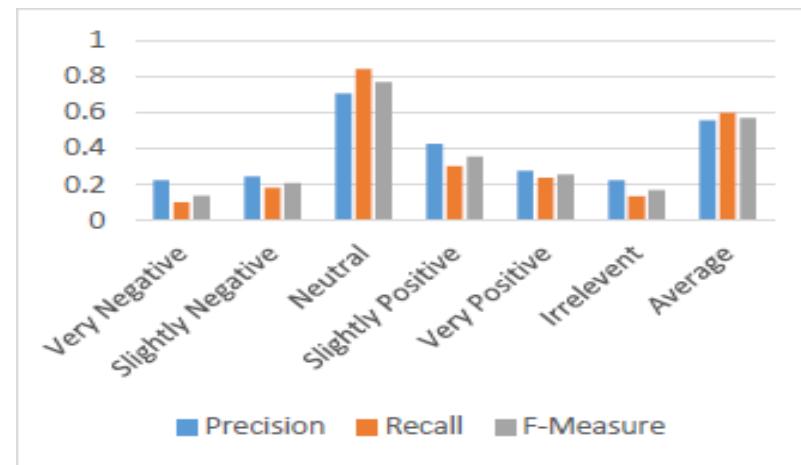
False Positives - liczebność fałszywych przypisań pozytywnego wydżwięku, czyli takich, w których metoda przypisała wydżwięk pozytywny do komentarzy ocenionych przez respondentów jako negatywne lub neutralne.

False Negatives - liczebność fałszywych przypisań negatywnego i neutralnego wydżwięku do komentarzy ocenionych przez respondentów jako pozytywne.

Twitter & SVM

Table 1. Twitter dataset for self-driving cars

Class	Tweets
Very Negative	110
Slightly Negative	685
Neutral	4245
Slightly Positive	1444
Very Positive	459
Irrelevant	213
Total	7156



Class	Precision	Recall	F-Measure
Very Negative	0.224	0.1	0.138
Slightly Negative	0.247	0.184	0.211
Neutral	0.708	0.841	0.769
Slightly Positive	0.428	0.305	0.356
Very Positive	0.278	0.237	0.256
Irrelevant	0.225	0.136	0.17
Average	0.558	0.599	0.572

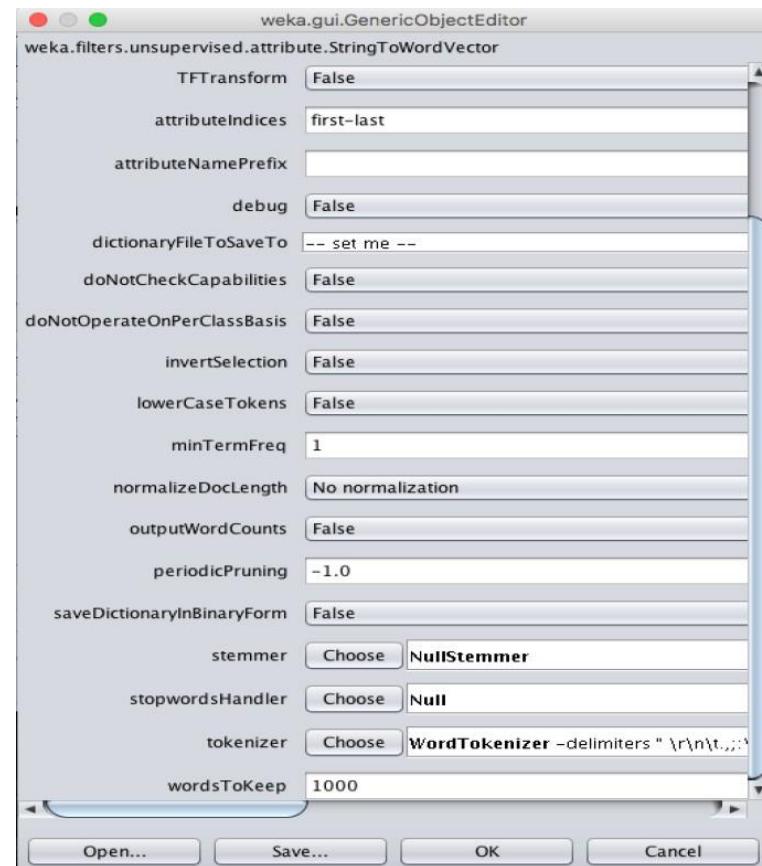
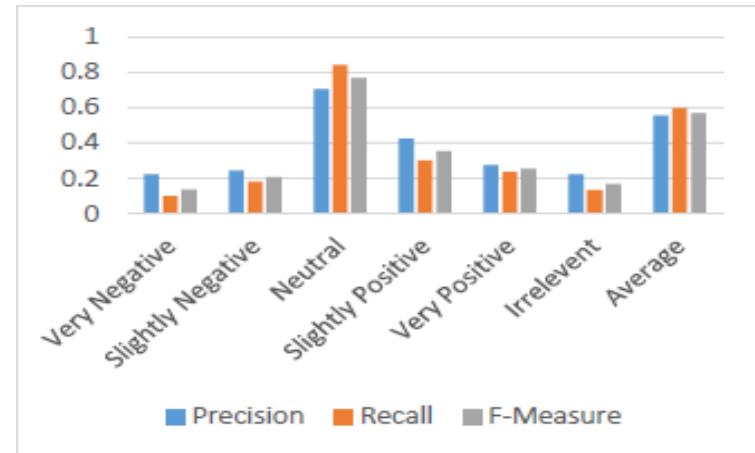
Ahmad, M., Aftab, S., & Ali, I. (2017). Sentiment analysis of tweets using svm. Int. J. Comput. Appl, 177(5), 25-29.

Twitter & SVM

Table 2. Twitter dataset for Apple products

Class	Tweets
Negative	1218
Neutral	2162
Positive	423
Irrelevant	81
Total	3884

Class	Precision	Recall	F-Measure
Negative	0.732	0.602	0.661
Neutral	0.729	0.859	0.789
Positive	0.548	0.376	0.446
Irrelative	0.318	0.173	0.224
Average	0.702	0.712	0.699



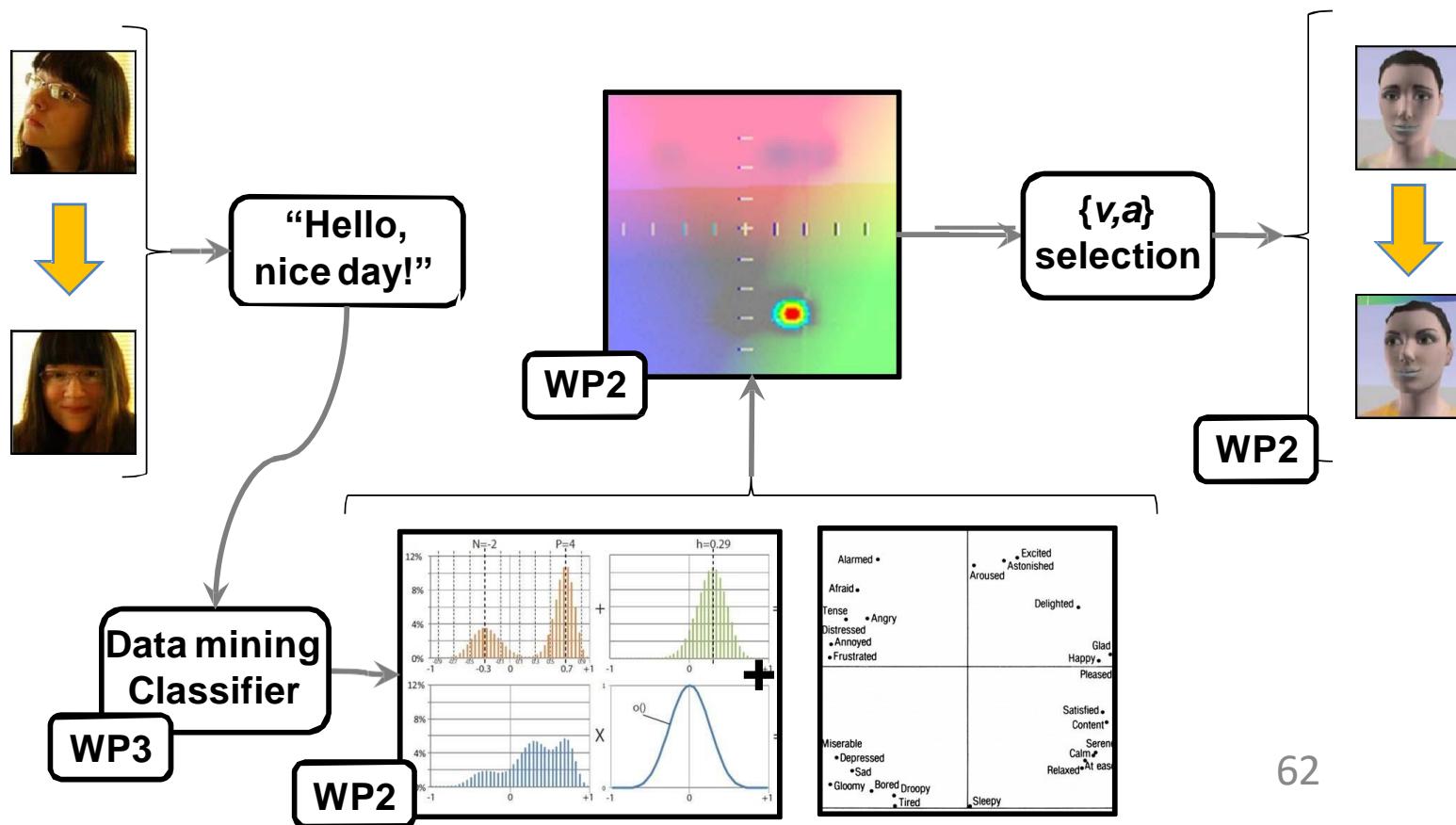
Inne rozwiązania afektywne

Projekt

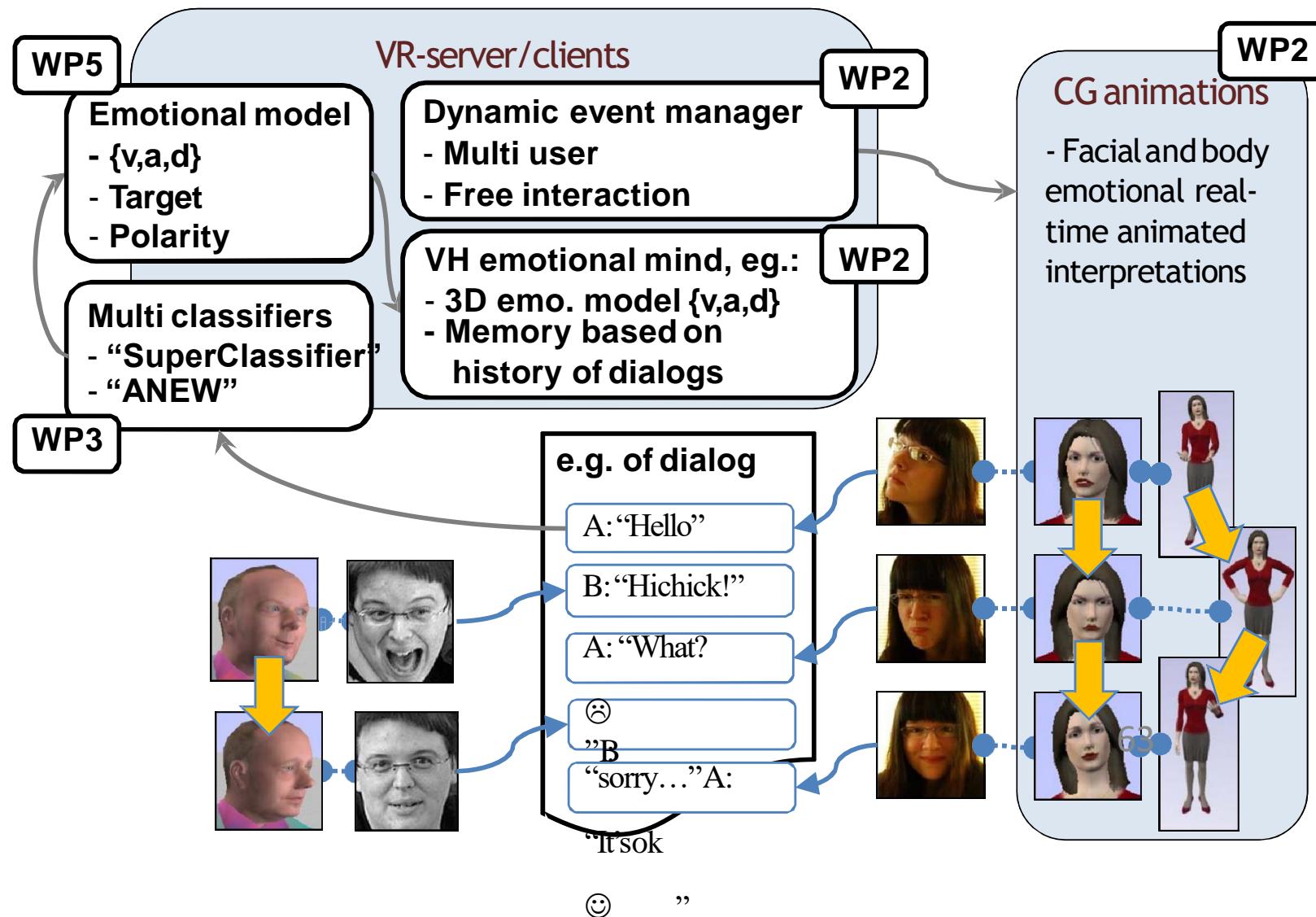


- Automatyczne gromadzenie i klasyfikacja danych o nastrojach w różnych e-społecznościach, a także wzajemna walidacja takich klasyfikatorów z wykorzystaniem metod psychofizjologicznych.
- Jakościowa i ilościowa analiza danych o sentymencjach oraz modelowanie oparte na danych zbiorowych emocji przez ABM, złożone sieci i paradygmaty skalowania wahań.
- Opracowanie inteligentnych emocjonalnie narzędzi ICT, takich jak afektywne systemy dialogu i graficznie animowane wirtualni agenci komunikujący się za pomocą interakcji emocjonalnych.

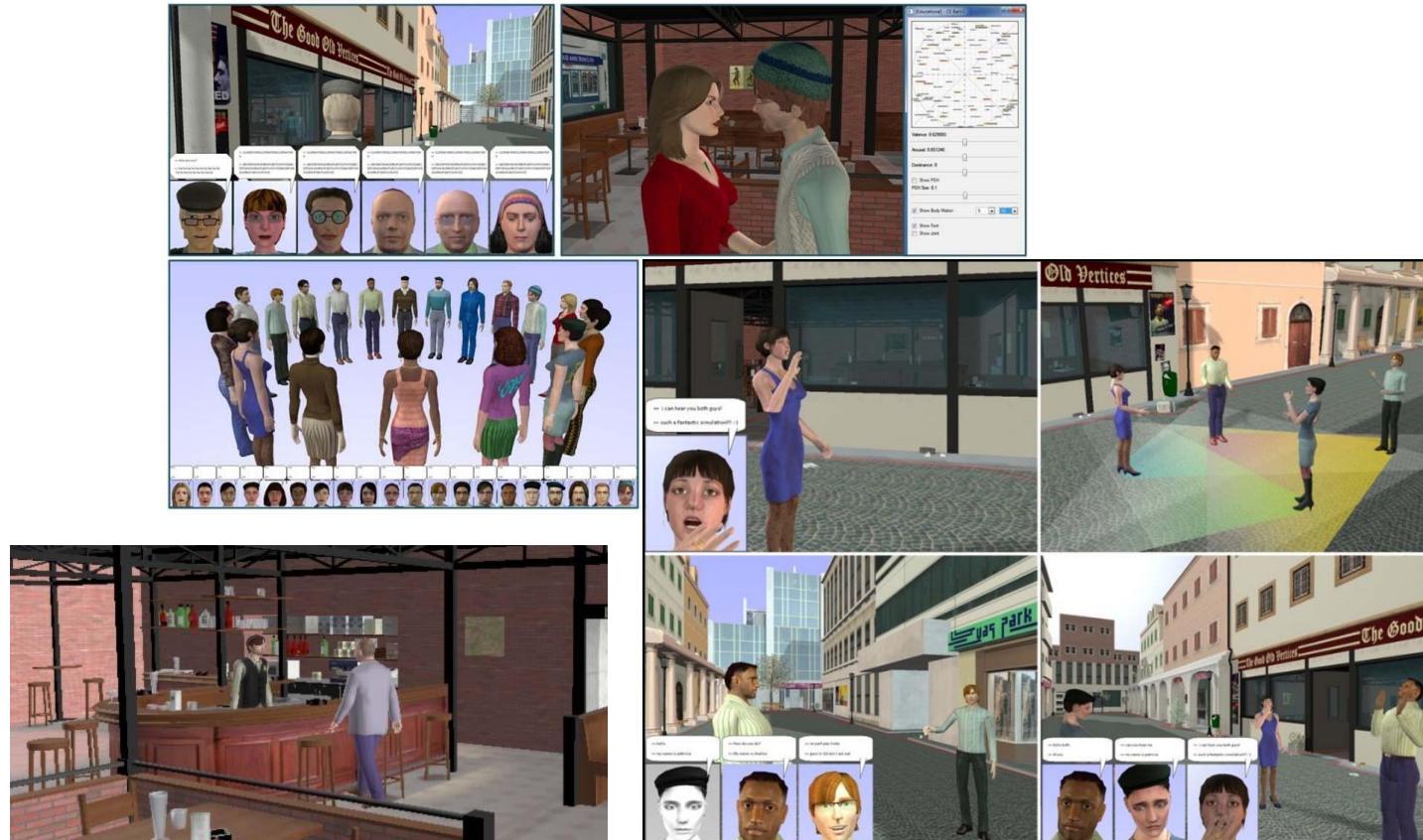
Detekcja emocji z głosu



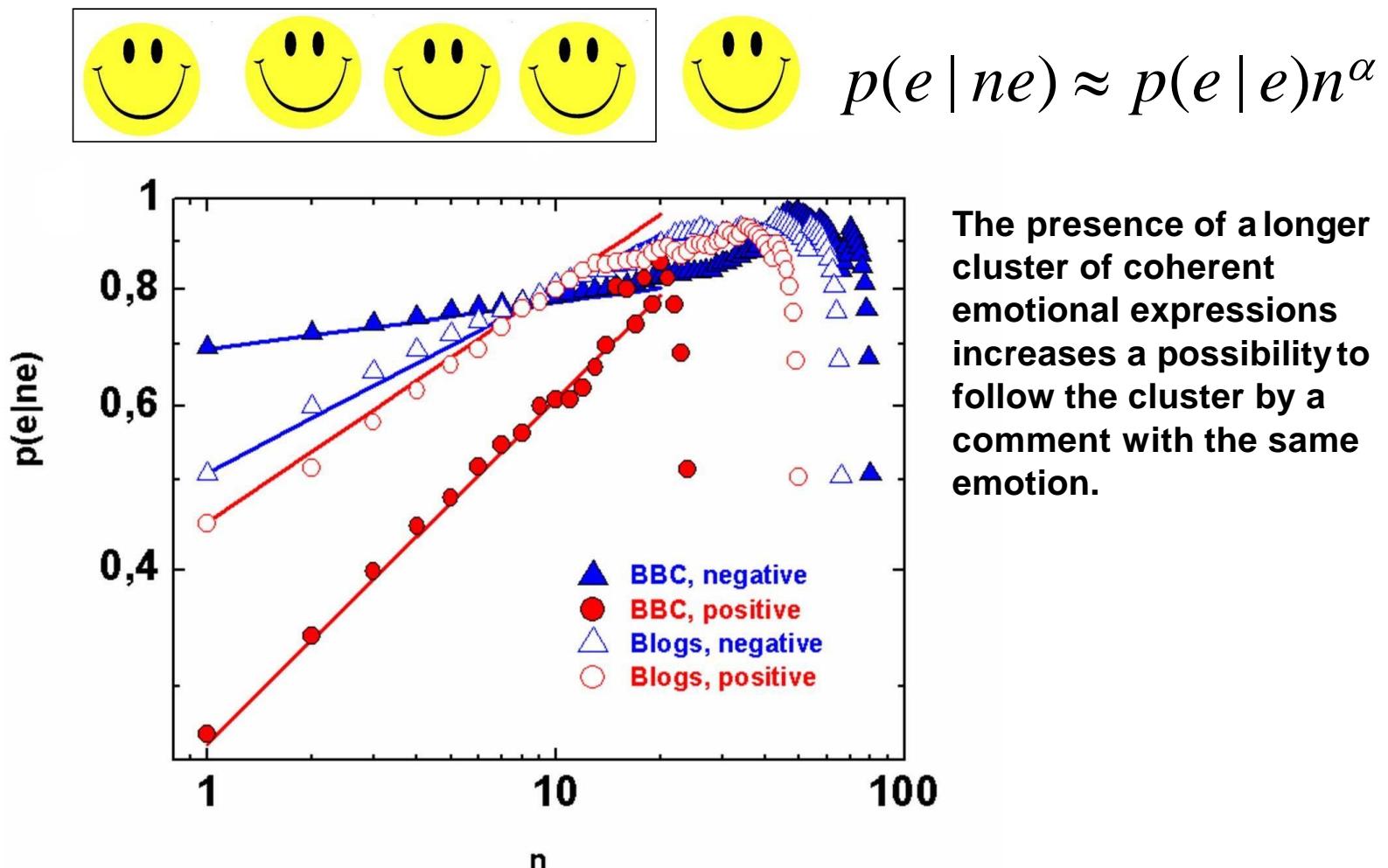
Systemy call center



Virtual Reality

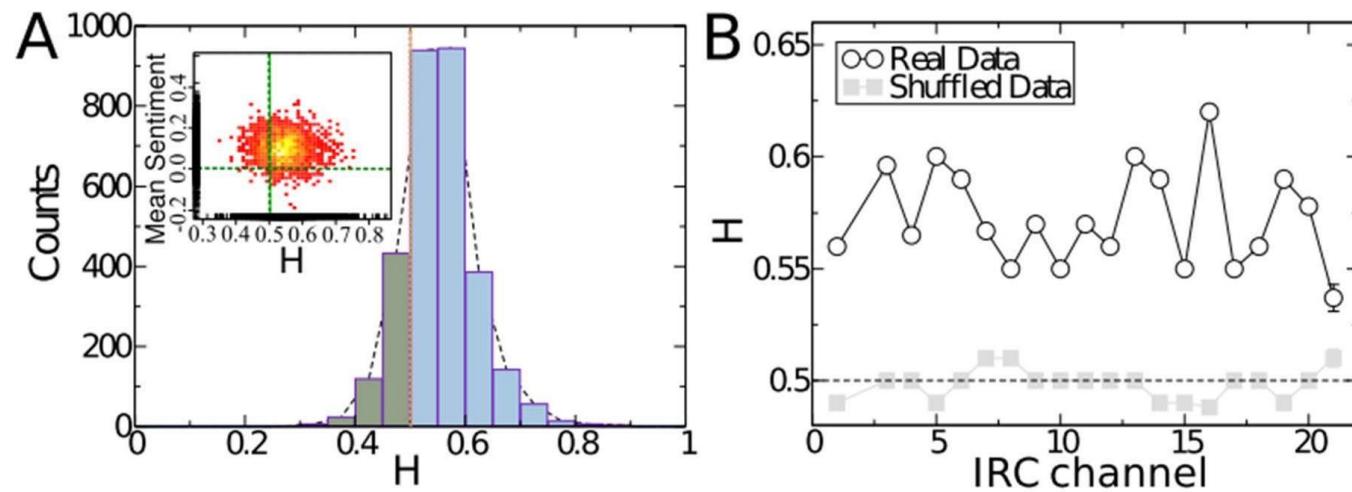


Homofilia & Emocje



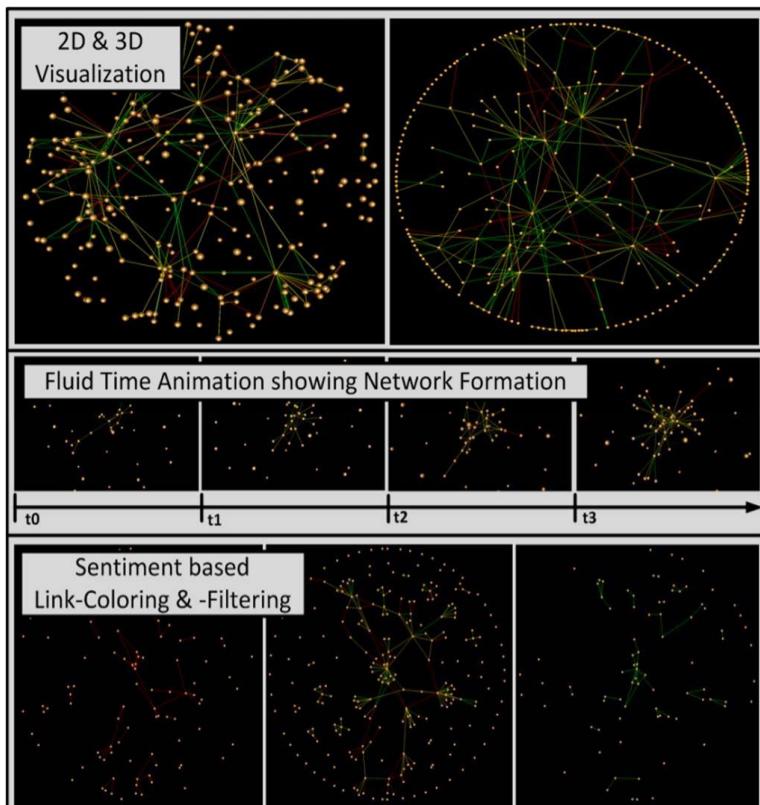
Conditional probability for cluster growth increases as a power-law with cluster length.

Systemy czatów & kanały IRC



- IRC chatroom user persistence in emotional expression.
- Individual persistence is independent of mean emotion.
- Persistence emerges in the conversation from user communication.

CMXViewer

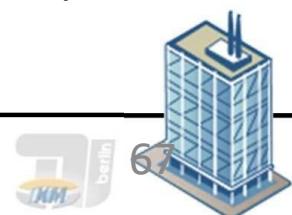


Workpackage 8
(TUB)

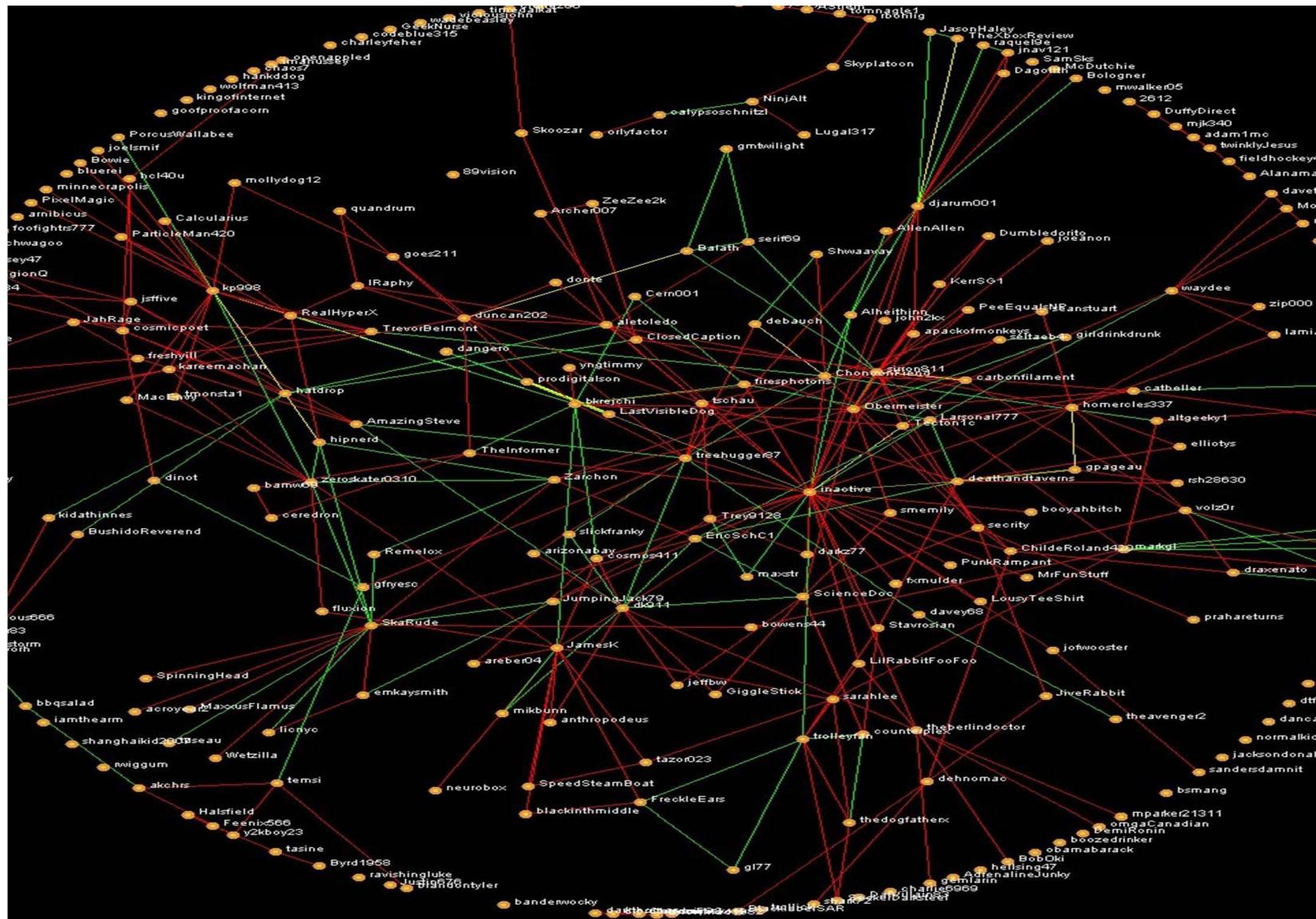


Development of Demonstrator
Software CMXViewer (D 8.2):

- Animated Graph Visualization representing sentiment propagation processes
- Dynamic visualization of longitudinal networkdata in 2D and 3D showing sentiment dissemination processes
- Sentiment-based dynamic link coloring

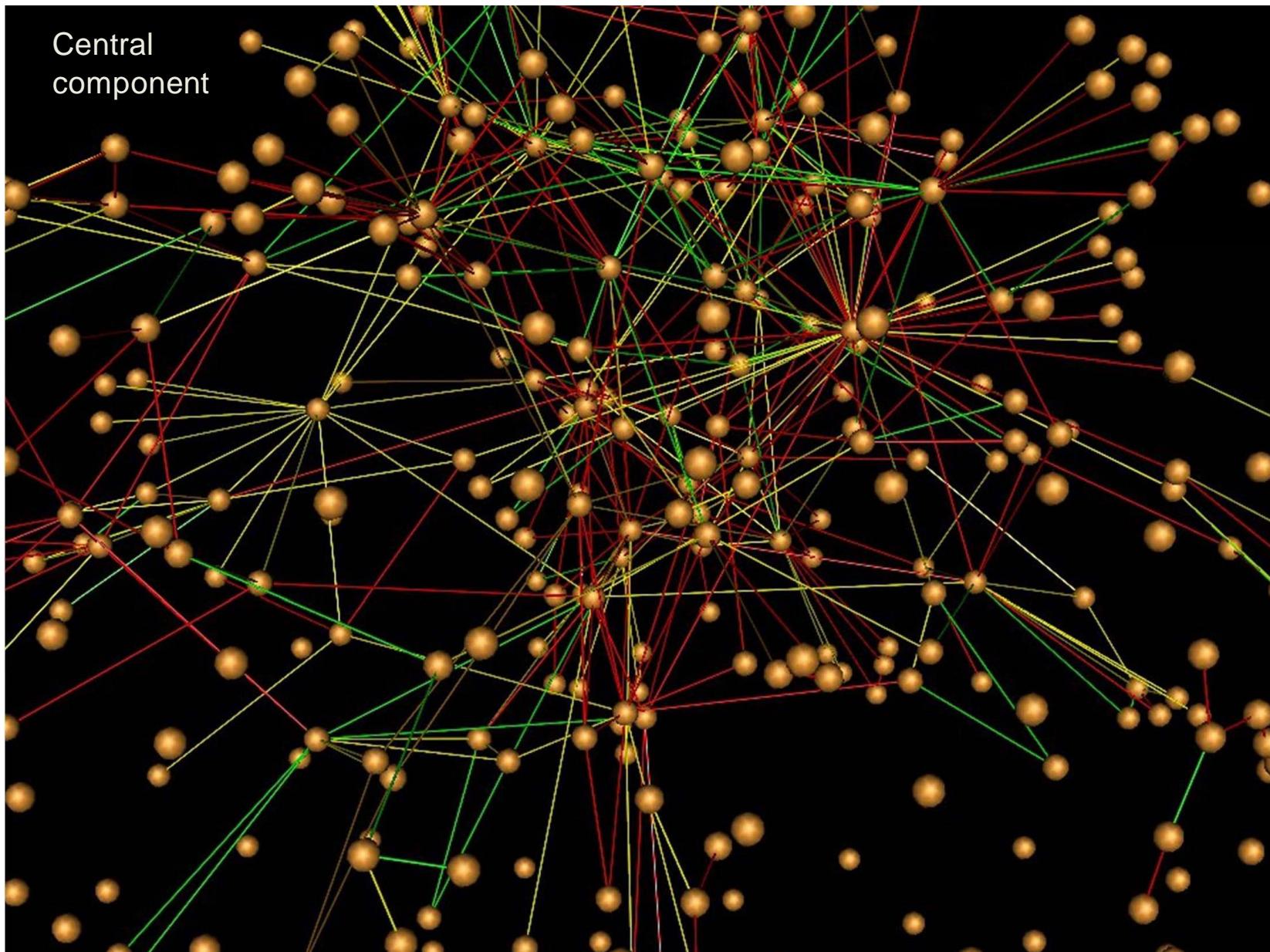


CMXViewer

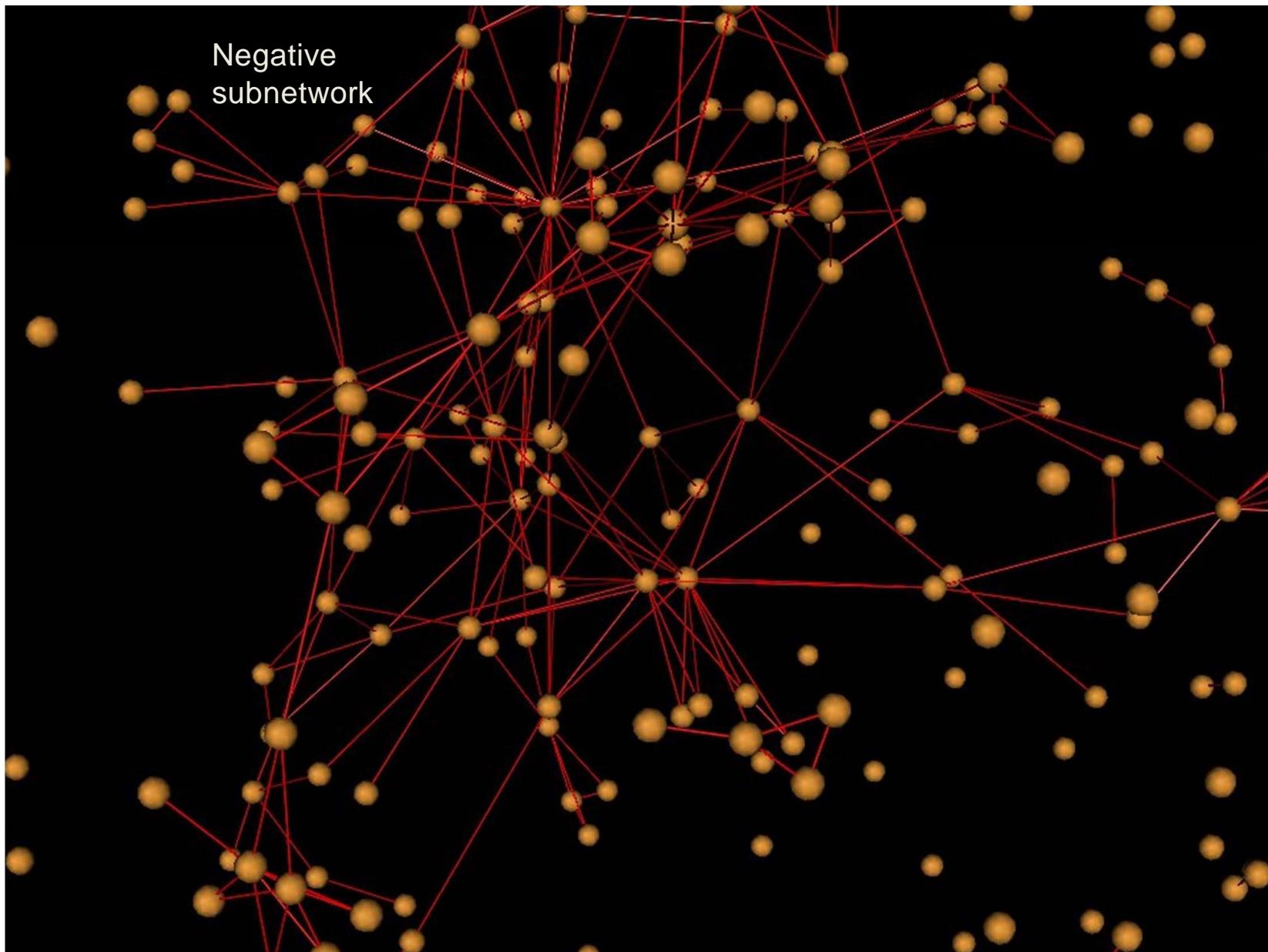


DIGG discussion network (Discussion ID 11223766) rendered with the CMXViewer(TUB)

CMXViewer



CMXViewer



CMXViewer



SensiStrength

[Test](#)- [Download](#) - [Java Version](#) - [Non-English](#) - [Buy!](#) - [About](#)



Automatic sentiment analysis of up to 16,000 social web texts per second with up to human level accuracy for English - other languages available or easily added.

SentiStrength estimates the *strength* of positive and negative sentiment in *short texts*, even for informal language. It has [human-level accuracy](#) for short social web texts in English, except political texts. SentiStrength reports *two* sentiment strengths:

-1 (not negative) to -5 (extremely negative)

1 (not positive) to 5 (extremely positive)

Why does it use two scores? Because [research from psychology](#) has revealed that we process positive and negative sentiment in parallel - hence mixed emotions.

SentiStrength can also report binary (positive/negative), trinary (positive/negative/neutral) and single scale (-4 to +4) results. SentiStrength was originally developed for English and optimised for general short social web texts but can be configured for other languages and contexts by changing its input files - some variants are demonstrated below.

SensiStrength

[Home](#) - [Test](#) - [Non-English](#) - [Download](#) - [Java Version](#) - [Buy](#) - [About](#)



The text 'I love you but hate the current political climate.'
has positive strength **3** and negative strength **-4**

*Approximate classification rationale: I love[3] you but hate[-4] the current political
climate .[sentence: 3,-4] [result: max + and - of any sentence][overall result = -1 as
pos<-neg] (Detect Sentiment)*

Positive sentiment strength ranges from 1 (not positive) to 5 (extremely positive) and negative sentiment strength from -1 (not negative) to -5 (extremely negative). The sentiment strength detection results are not always accurate - they are guesses using a set of rules to identify words and language patterns usually associated with sentiment.

Another Go? Try a [non-English experimental version](#)?

Enter text:

Keyword test: Specify keywords for the sentiment classification

Enter keywords (comma-separated list, no spaces: exact matches only -e.g., add
mike,mike's,mikes if you want to match all variants):

Mozdeh

Mozdeh - Big Data Text Analysis - [covid: 179 matches found out of 1134 (15,8%) (word ID: 6) Av pos:1,3408; av neg:2,0168]

File Analyse Networks Data Subprojects Advanced About

Advanced Search

covid
OR is default: Use AND between words to match all words; AND/OR priority /IGNORED: processing is left to right

Search Next Page Show of Project: window Previous Sort by: Date (ascending)

Advanced Search Tips

Report a maximum of one match per user per interval ID of first item to show 2020.4.16 First date to show 38 Results per page Hide #tags and @s Hide duplicates (v. slow)

Only return results from this topic label authors (case sensitive)

Results only from this user ID Only posted by label user Select alternative subproject

User gender Text must be in username / URL Exclude matching URL/source instead Exclude users with this many posts collected: 0 ≤ retweets ≤ 165

Positive sentiment Negative sentiment

Proportion of texts

Negative sentiment

Positive sentiment

Association mining options

Report words occurring at least this often: 5 Report words occurring at most this often: 50000 Words to show: 1000 List by relative frequency List by difference in proportions Z (-ve if in fewer) List by chisquare (signed)

Load Word Freq List Reference Set

Twitter texts

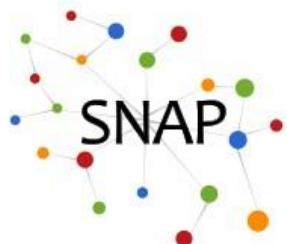
Publication date: 15.04.2020, 15.04.2020, 16.04.2020, 16.04.2020, 17.04.2020, 15.04.2020, 15.04.2020, 16.04.2020, 16.04.2020, 17.04.2020

Words most associating with the current search and filters

Word	MatchPc	NoMatch	Matches	Total	DiffPZ	ChiSq	Sig (2202 tests)
covid	21,2%	14,8%	38	179	2,2	4,7	covid
mamy	2,2%	1,2%	4	15	1,2	0,7	
żeby	1,7%	0,8%	3	11	1,1	0,4	
@morawieckim	1,7%	0,8%	3	11	1,1	0,4	
śmierietność	1,7%	0,9%	3	12	0,9	0,2	
takie	1,7%	1,0%	3	13	0,7	0,1	
sobie	2,2%	1,6%	4	19	0,6	0,1	
@premierpp	1,1%	1,0%	2	12	0,1	0,1	
mają	1,7%	1,3%	3	15	0,5	0,0	
samo	1,1%	0,8%	2	10	0,4	0,0	
pierwsze	1,1%	0,8%	2	10	0,4	0,0	
tu	1,7%	1,4%	3	16	0,3	0,0	
ludzie	1,1%	1,5%	2	16	-0,4	0,0	
gdzie	1,1%	1,5%	2	16	-0,4	0,0	
virusem	0,6%	0,9%	1	10	-0,5	0,0	
tyle	0,6%	0,9%	1	10	-0,5	0,0	
jeżeli	0,6%	0,9%	1	10	-0,5	0,0	
prof	1,1%	1,4%	2	15	-0,3	0,0	
we	1,1%	1,4%	2	15	-0,3	0,0	
że	3,9%	4,1%	7	46	-0,1	0,0	

/4

Zbiory danych



SNAP for C++ ▶
SNAP for Python ▶
SNAP Datasets ▶
BIOSNAP Datasets
What's new
People
Papers
Projects ▶
Citing SNAP
Links
About
Contact us

Open positions

Open research positions in SNAP group are available at undergraduate, graduate and postdoctoral levels.

• Web data: Amazon reviews

• Dataset information

This dataset consists of reviews from [amazon](#). The data span a period of 18 years, including ~35 million reviews up to March 2013. Reviews include product and user information, ratings, and a plaintext review. **Note:** this dataset contains potential duplicates, due to products whose reviews Amazon merges. A file has been added below ([possible_dupes.txt.gz](#)) to help identify products that are potentially duplicates of each other.

Note: A new-and-improved Amazon dataset is available [here](#), which corrects the above duplication issues, and also contains more complete data/metadata.

Dataset statistics

Number of reviews	34,686,770
Number of users	6,643,669
Number of products	2,441,053
Users with > 50 reviews	56,772
Median no. of words per review	82
Timespan	Jun 1995 - Mar 2013

• Source (citation)

- J. McAuley and J. Leskovec. [Hidden factors and hidden topics: understanding rating dimensions with review text](#). RecSys, 2013.

• Files

File	Description	Size
------	-------------	------

<https://snap.stanford.edu/data/web-Amazon.html>

<https://blog.cambridgespark.com/50-free-machine-learning-datasets-sentiment-analysis-b9388f79c124>

Dziękuję za uwagę