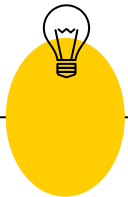


Sistem Temu Kembali Informasi

“Opinion Mining (Sentiment Analysis)”





Buku Penunjang & Literatur





Pengantar

- Dua jenis utama informasi tekstual.
 - Fakta dan Pendapat (Opini)
 - Catatan: pernyataan faktual dapat menyiratkan pendapat juga.
- Sebagian besar metode pemrosesan informasi teks terkini (misalnya, pencarian web (web search) , penambangan teks (text mining) bekerja dengan informasi faktual.
- Analisis sentimen atau Opinion Mining
 - studi komputasional tentang pendapat, sentimen dan emosi yang diungkapkan dalam teks.
- Mengapa perlu Opinion Mining (Sentiment Analysis)? Terutama karena Web; jumlahnya komentar teks yang sangat besar.



Pendahuluan - media buatan pengguna

● Pentingnya Opini:

- Opini itu penting karena kapan pun kita perlu membuat keputusan, kita ingin mendengar opini orang lain.
- Di masa lalu,
 - Individu: opini dari teman dan keluarga
 - bisnis: survei, kelompok fokus, konsultan ...

● Dari mulut ke mulut di Web

- Media yang dibuat pengguna: Seseorang dapat mengungkapkan opini tentang apa pun dalam ulasan, forum, kelompok diskusi, blog ...
- Pendapat skala global: Tidak lagi terbatas pada:
 - Individu: lingkaran teman seseorang
 - Bisnis: Survei berskala kecil, kelompok fokus kecil, dll.



Masalah yang Menarik!

☉ Intelektual menantang & aplikasi utama.

- Topik penelitian yang populer dalam beberapa tahun terakhir di NLP dan data mining Web.
- 20-60 perusahaan di Amerika Serikat saja

☉ Ini menyentuh setiap aspek NLP namun dibatasi.

- Sedikit penelitian di NLP / Linguistik di masa lalu.

☉ Berpotensi teknologi utama dari NLP.

- Tapi "belum" dan tidak mudah!
- Sumber data dan integrasi data juga sulit!



Contoh Review

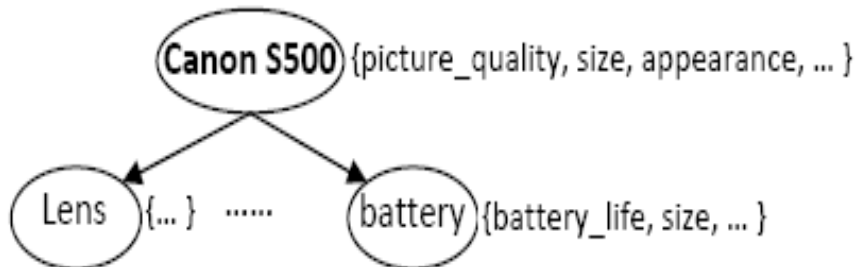
- ☉ *"Saya membeli iPhone beberapa hari yang lalu. Telepon itu bagus sekali. Layar sentuhnya sangat keren. Kualitas suaranya pun jernih. Meski masa pakai baterai tidak lama, tidak apa-apa buat saya. Namun, ibuku marah padaku karena aku tidak memberitahunya sebelum membeli telepon. Dia juga mengira teleponnya terlalu mahal, dan ingin aku mengembalikannya ke toko. ... "*
- ☉ Apa yang kita lihat?

Opini, **target opini**, dan **pemegang opini**



Objek Sasaran (Liu, buku Web Data Mining, 2006)

- **Definisi (object):** Sebuah *objek o* adalah produk, orang, event, organisasi, atau topik. *o* direpresentasikan sebagai:
 - hirarki **komponen**, **sub komponen**, dan sebagainya.
 - Setiap node mewakili komponen dan dikaitkan dengan satu set **atribut** dari komponennya



- Suatu pendapat dapat diekspresikan pada setiap node atau atribut dari node.
- Untuk menyederhanakan diskusi kita, kita menggunakan *fitur* istilah untuk mewakili kedua komponen dan atribut.



Apa itu Opini? (Liu, a Ch dalam buku pegangan NLP)

☉ Sebuah *opini* adalah quintuple $\Rightarrow (o_j, f_{jk}, so_{ijkl}, h_i, t_l)$,

where

- o_j adalah objek target.
- f_{jk} adalah fitur dari objek terget.
- so_{ijkl} adalah nilai sentimen dari pemegang opini (h_i) pada fitur objek targetnya.
- h_i adalah pemegang opini.
- t_l adalah waktu ketika opini diungkapkan.



Tujuan - struktur yang tidak terstruktur

- **Tujuan:** Diberikan dokumen berpendirian,
 - Temukan semua quintuples (*oj, fjk, soijkl, hi, tl*),
 - yaitu, saya mengumpulkan lima informasi yang sesuai di setiap kuintuple, dan
 - Atau, selesaikan beberapa masalah yang lebih sederhana
- Dengan kuintuples,
 - Teks Tidak Terstruktur → Data Terstruktur
 - Alat data dan visualisasi tradisional dapat digunakan untuk mengiris, mencocokkan dan memvisualisasikan hasilnya dengan berbagai cara
 - Mengaktifkan analisis kualitatif dan kuantitatif.



Sentimen Klasifikasi: doc-level (Pang dan Lee, et al 2002 dan Turney 2002)

- **Klasifikasikan sebuah dokumen** (misalnya, ulasan) berdasarkan keseluruhan sentimen yang diungkapkan oleh pemegang pendapat
 - **Kelas**: Positif, atau negatif (dan netral)
 - Dalam model, $(o_j, f_{jk}, so_{ijkl}, h_i, t_l)$,
 - **Ini mengasumsikan**
 - Setiap dokumen berfokus pada satu objek dan berisi opini dari satu pemegang pendapat.
 - Ini mempertimbangkan pendapat tentang objek, o_j (atau , o_j (or $o_j = f_{jk}$)



Analisis Subjektivitas (Wiebe et al, 2004)

- Analisis sentimen tingkat sentimen memiliki dua tugas:
 - **Klasifikasi Subjek:** Subyektif atau objektif.
 - **Tujuan:** misalnya, *saya membeli iPhone beberapa hari yang lalu.*
 - **Subyektif:** misalnya, *Hal ini seperti telepon bagus.*
 - **Klasifikasi sentimen:** Untuk kalimat subjektif atau klausa, klasifikasikan positif atau negatif.
 - **Positif:** *Itu adalah ponsel yang bagus.*
- **Namun.** (Liu, Bab dalam buku pegangan NLP)
 - kalimat subyektif ≠ ve atau -ve pendapat
 - Misalnya, *saya pikir dia datang kemarin.*
 - Kalimat obyektif ≠ tidak ada pendapat
 - **Menyiratkan pendapat ve:** *Ponsel saya pecah di hari kedua.*



Analisis Sentimen Berbasis-Fitur

- ◎ Klasifikasi sentimen pada tingkat dokumen dan kalimat (atau klausa) **tidak mencukupi**,
 - mereka tidak tahu apa yang orang suka dan / atau tidak suka
 - Pendapat positif pada suatu objek tidak berarti bahwa pemegang pendapat menyukai semuanya.
 - Pendapat negatif pada suatu objek tidak berarti
- ◎ Tujuan: **Menemukan semua kuintuples**
$$(o_j, f_{jk}, so_{ijkl}, h_i, t_l)$$
- ◎ Dengan semua kuintuples, semua jenis analisis menjadi mungkin.



Ringkasan Opini Berbasis Fitur (Hu & Liu, KDD-2004)

- ☉ "Saya membeli iPhone beberapa hari yang lalu. Telepon itu bagus sekali. *Layar sentuhnya* sangat keren. *Kualitas suaranya* pun jernih. Meski masa *pakai daya tahan baterai* tidak lama, tidak apa-apa buat saya. Namun, ibuku marah padaku karena aku tidak memberitahunya sebelum membeli telepon. Dia juga mengira teleponnya *terlalu mahal*, dan ingin aku mengembalikannya ke toko. ... "

Ringkasan Fitur Berdasarkan:

Feature1: **Layar Sentuh**

Positive: 212

- *Layar sentuh* benar-benar keren.
- *Layar sentuh* begitu mudah digunakan dan bisa melakukan hal yang menakjubkan.

Negative: 6

- *Layarnya* mudah tergores.
- Saya memiliki banyak kesulitan dalam melepaskan bekas jari dari *layar sentuh*.

Feature2: **battery life**

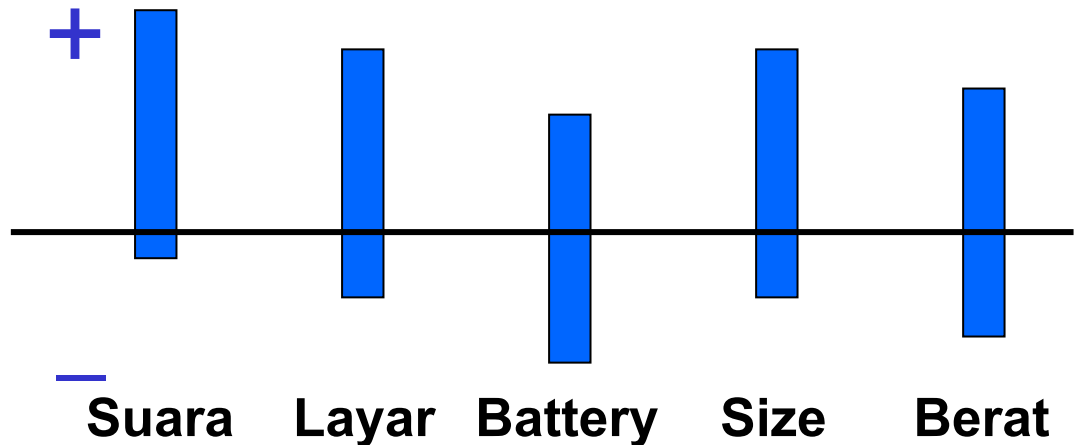
...

Note: We omit opinion holders

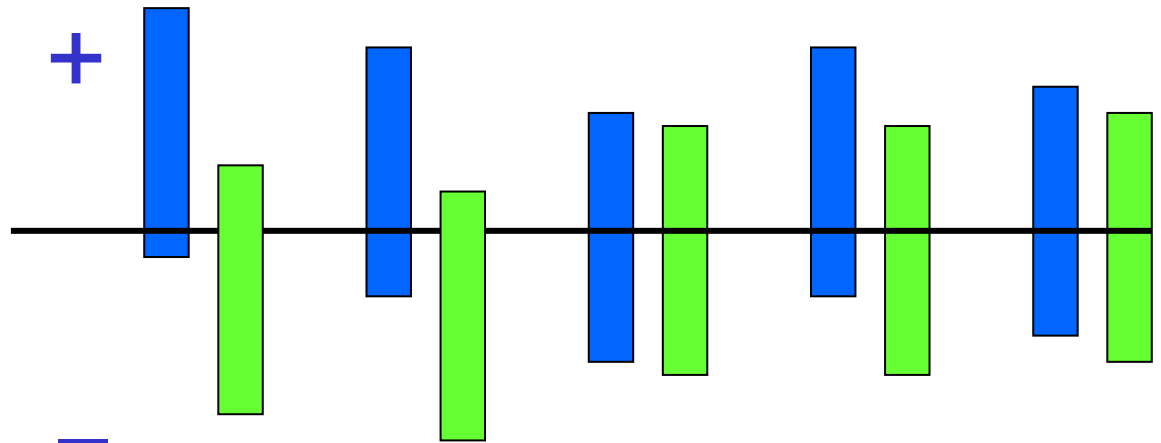
Perbandingan Visual (Liu dkk. WWW-2005)



- Ringkasan review dari Cell Phone 1



- Perbandingan review dari Cell Phone 1
- Cell Phone 2



Demo Langsung: OpinionEQ



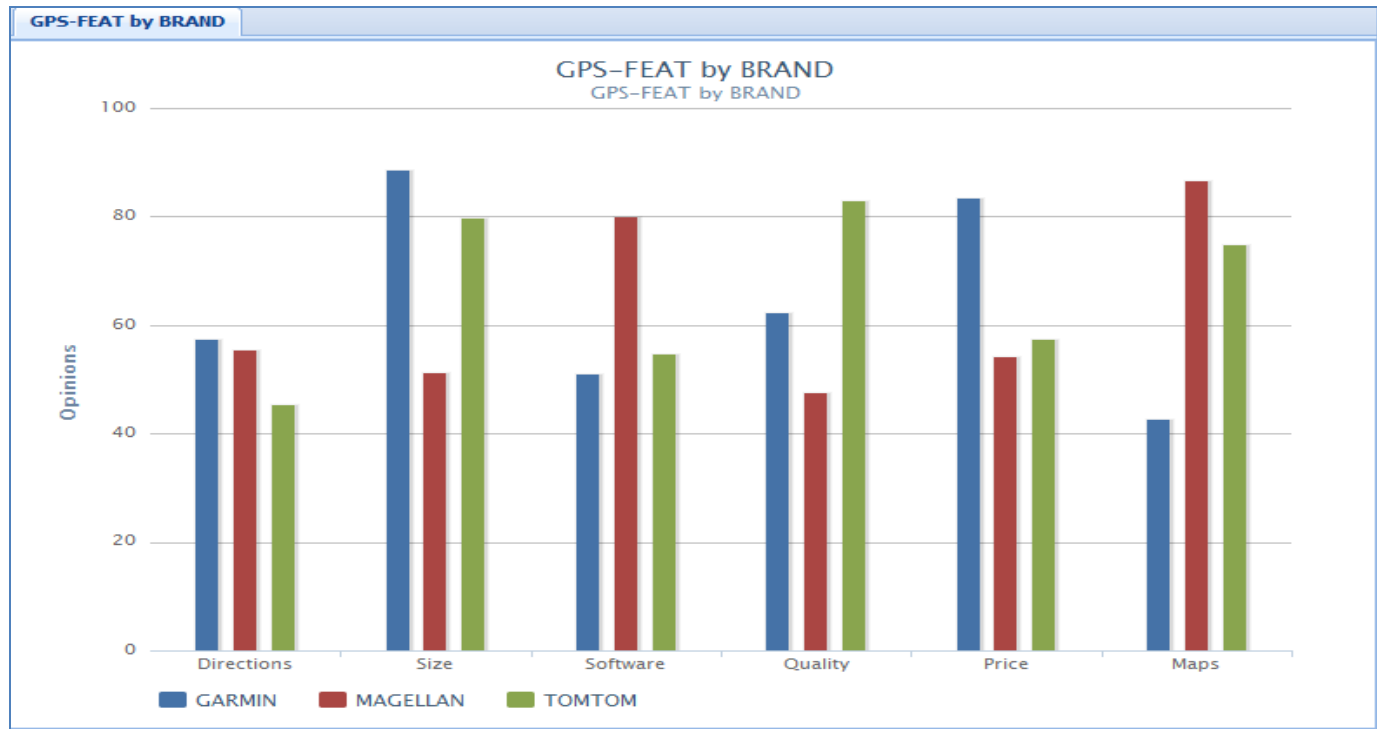
(saya memberikan demo langsung sistem OpinionEQ. Beberapa screensdumps dari demo ditampilkan di sini)

- **Demo 1:** Bandingkan pendapat konsumen mengenai tiga sistem GPS, **Garmin, Magellan, Tomtom.**
 - Berdasarkan seperangkat fitur, **harga, peta, perangkat lunak, kualitas, ukuran**, dll.
- **Demo 2:** Analisis halaman instan
 - Pengguna memberikan URL, dan sistem tersebut mengidentifikasi opini di halaman secara instan.
- Kami juga memiliki sistem pemantauan opini Twitter (tidak demo-red)



Demo 1: Bandingkan 3 GSP pada fitur yang berbeda

- Setiap bar menunjukkan proporsi + ve pendapat





Demo 1: Rincian pendapat kalimat

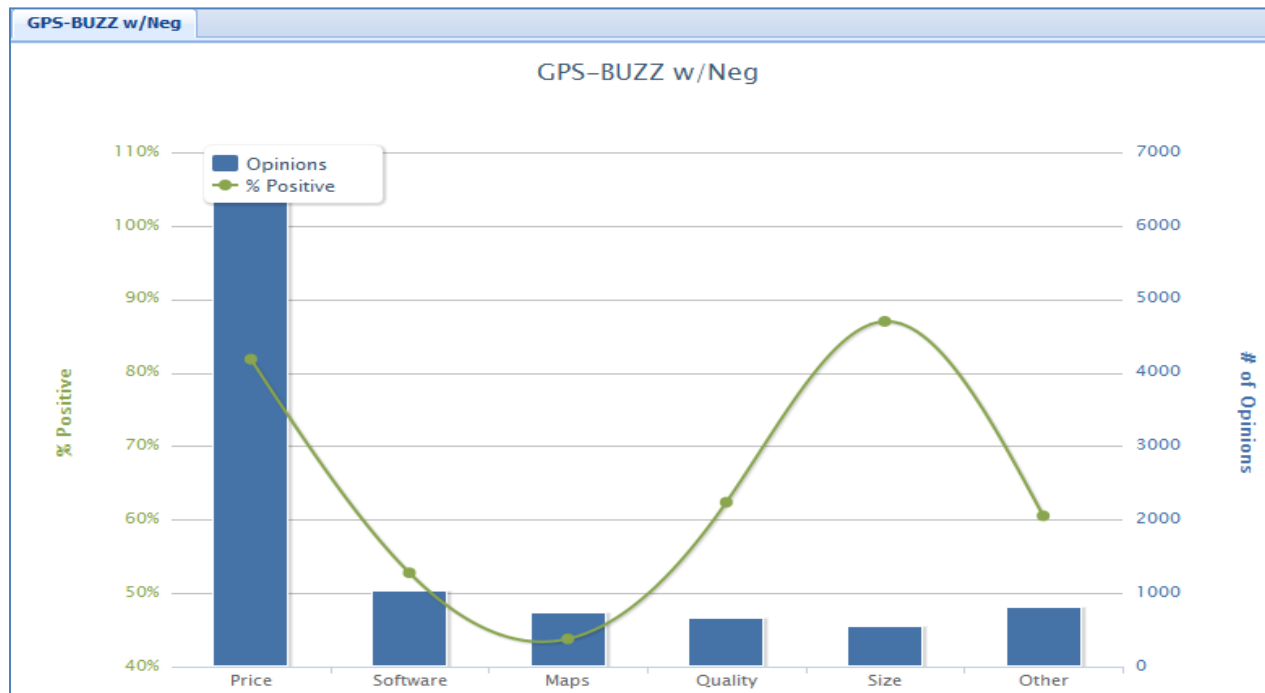
- Anda bisa klik pada setiap bar untuk melihat kalimat opini. Berikut adalah kalimat opini negatif pada fitur peta Garmin.
- Diagram lingkaran memberi proporsi pendapat.





Demo 1: # dari fitur menyebutkan

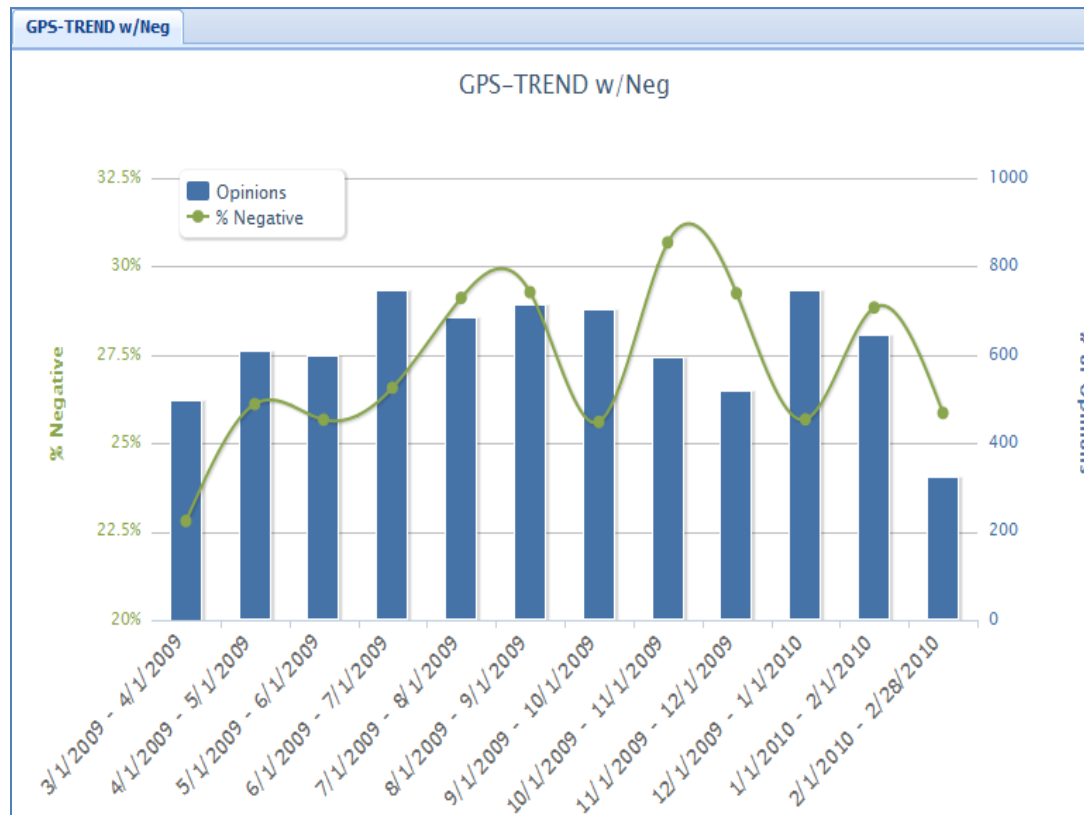
- Orang banyak berbicara tentang harga dibanding fitur lainnya. Mereka cukup positif tentang harga, tapi bukan tentang peta dan perangkat lunak.





Demo 1: Tren opini agregat

- Lebih mengeluh pada bulan Juli - Agustus, dan di bulan Oktober - Dec!





Barang lain dari OpinionEQ

- Izinkan pengguna untuk memilih
 - Produk / merek,
 - fitur
 - Situs
 - Periode waktu
- **untuk perbandingan pendapat**
- Bekerja pada fitur individu untuk analisis terperinci.
- Izinkan pengguna melihat teks opini lengkap dan juga laman sebenarnya di situs dari tempat teks opini diekstrak.



Demo 2 - Instant page analysis

- Dengan URL, secara otomatis mengidentifikasi opini di laman. Hijau: + ve, dan merah: -ve

The screenshot shows the Amazon.com interface for a product page. The top navigation bar includes the Amazon logo, a sign-in prompt, and links for 'Your Amazon.com', 'Today's Deals', 'Gifts & Wish Lists', and 'Gift Cards'. Below this is a search bar with 'Watches' entered. The main content area is titled 'Customer Reviews' for the 'Casio P015-1K Travel Alarm Clock with Thermometer'. It displays a star rating distribution: 5 stars (23), 4 stars (8), 3 stars (9), 2 stars (1), and 1 star (7). The average customer review is 4.5 stars. A 'Create your own review' button is present. To the right, there is a 'Search Customer Reviews' box with a 'GO!' button and a checkbox for 'Only search this product's reviews'. Below the reviews section, there is a comparison of 'The most helpful favorable review' and 'The most helpful critical review'. The favorable review is 5 stars and describes the clock as dependable and durable. The critical review is 3 stars and mentions a failed alarm switch. A 'Vs.' icon is placed between the two reviews. At the bottom of each review section, there are links to see more reviews of a specific star rating.

amazon.com Hello, [Sign in](#) to get personalized recommendations. New customer? [Start here](#).

Your Amazon.com Today's Deals [Gifts & Wish Lists](#) [Gift Cards](#)

Shop All Departments Search

Watches What's New Men's Watches Women's Watches Kids' Watches Accessories

Customer Reviews

[Casio P015-1K Travel Alarm Clock with Thermometer](#)

48 Reviews

Average Customer Review (48 customer reviews)

Share your thoughts with other customers [Create your own review](#)

Search Customer Reviews [GO!](#)

☒ Only search this product's reviews

The most helpful favorable review	The most helpful critical review
<p>27 of 27 people found the following review helpful:</p> <p> great little clock</p> <p>I've owned this clock for several years - it's dependable, durable and reliable. I find it great to have the temperature readout, in both Fahrenheit and Celsius (that way I can whine bi-lingually when it's hot..). I've travelled with it, camped with it, and use it as my everyday bedside alarm. My only complaint is that the snooze button is set for only 5 minutes per...</p> <p>Read the full review ></p> <p>Published on June 7, 2006 by USA-Canada</p> <p>> See more 5 star, 4 star reviews</p>	<p>62 of 69 people found the following review helpful:</p> <p> A missed target for CASIO</p> <p>I've owned a CASIO PQ-10 Travel Alarm Clock for about ten years, traveled with it thousands of miles in my bicycle panniers, and it's still running on the original batteries! The alarm switch finally failed, and I really loved this clock, so I decided to try the PQ-15 as a replacement since the PQ-10 is no longer available. I'm dissapointed - the PQ-15 is more than twice...</p> <p>Read the full review ></p> <p>Published on September 22, 2006 by A. Pietsch</p> <p>> See more 3 star, 2 star, 1 star reviews</p>



Demo 2 - Instant page analysis

- Ini juga mengekstrak pendapat di halaman dan daftar mereka.

Pros	Cons
<p>1. FREE 2-Day Shipping: See details. . (details)</p> <p>2. great little clock. . (details)</p> <p>3. I've owned this clock for several years - it's dependable, durable and reliable. (details)</p> <p>4. I find it great to have the temperature readout, in both Fahrenheit and Celcius (that way I can whine bi-lingually when it's hot. . (details)</p> <p>5. The alarm switch finally failed, and I really loved this clock, so I decided to try the PQ-15 as a replacement since the PQ-10 is no longer available. . (details)</p> <p>6. love it!, May 10, 2010. . (details)</p> <p>7. While the digits of the display are decent enough, the operations to set the clock for time, alarm or any other of the settings, are horrific. (details)</p> <p>8. LLBean has a much nicer, user friendly travel clock for less and I could kick myself for not having taken the time to buy another of those.. . (details)</p> <p>9. I bought from Amazon due to the ease of billing, which was the ONLY reason I bought this instead of the LLBean clock.. . (details)</p> <p>10. Very accurate timing but temperture is off, April 9, 2010. (details)</p>	<p>1. I find it great to have the temperature readout, in both Fahrenheit and Celcius (that way I can whine bi-lingually when it's hot. . (details)</p> <p>2. My only complaint is that the snooze button is set for only 5 minutes per. . (details)</p> <p>3. A missed target for CASIO. . (details)</p> <p>4. The alarm switch finally failed, and I really loved this clock, so I decided to try the PQ-15 as a replacement since the PQ-10 is no longer available. . (details)</p> <p>5. I'm dissapointed - the PQ-15 is more than twice. . (details)</p> <p>6. I always seem to loose track of time on the computer and phone and needed something.. . (details)</p> <p>7. Not worth my time or money to return the clock, and I DO NOT recommend this clock at all.. . (details)</p> <p>8. LLBean has a much nicer, user friendly travel clock for less and I could kick myself for not having taken the time to buy another of those.. . (details)</p> <p>9. While you can't please everyone with a product, I would not recommend this to anyone.. . (details)</p> <p>10. Bought it in December 1996 wholesale for \$13 (more than 13 years ago)and i'm still using it EVERYDAY.. Bottom line: Very Very Very accurate and durable BUT temperture show</p>



Analisis Sentimen Menantang!

- “Terakhir ini Sabtu, saya membeli ponsel **Nokia** dan pacar saya membeli ponsel **Motorola** dengan **Bluetooth**. Kami saling menelepon saat sampai di rumah. **Suara di telepon saya tidak begitu jelas, lebih buruk dari telepon saya sebelumnya. Masa pakai baterai sudah lama. Pacar saya cukup senang dengan teleponnya. Saya ingin telepon dengan kualitas suara bagus. Jadi pembelian saya benar-benar mengecewakan. Saya mengembalikan telepon kemarin.**”



Sentimen Analisis bukan hanya satu masalah

- $(o_j, f_{jk}, so_{ijkl}, h_i, t_l),$
 - o_j - a target object: Dinamakan Entity Ekstraksi (lebih)
 - f_{jk} - a feature of o_j : Informasi Ekstraksi
 - so_{ijkl} adalah sentimen: Sentimen tekad
 - h_i adalah pemegang pendapat: Informasi / Data Ekstraksi
 - t_l adalah waktu: Data Ekstraksi
- Resolusi bersama referensi
- Ekstraksi relasi
- Pertandingan sinonim (suara = kualitas suara) ..
- Tak satu pun dari mereka adalah masalah yang dipecahkan!



Ekstraksi objek yang bersaing

- Pengguna pertama memberi beberapa objek / produk sebagai **bibit**, misalnya BMW dan Ford.
- Sistem kemudian mengidentifikasi objek pesaing lainnya dari corpus opini.
- Masalahnya bisa diatasi dengan **pembelajaran PU** (Belajar dari contoh positif dan tidak berlabel) (Liu et al, 2002, 2003).
- Lihat (Li et al. ACL-2010)



Ekstraksi fitur

- Kami mengusulkan pendekatan *propagasi ganda* dalam (Qiu et al. IJCAI-2009).
- Ini memanfaatkan hubungan ketergantungan opini dan fitur untuk mengekstrak fitur.
 - Pendapat kata-kata memodifikasi fitur objek, misalnya,
 - “Kamera ini mengambil *gambar yang bagus*”
- Bootstraps algoritma menggunakan seperangkat kata pendapat benih (tidak ada input fitur).
 - Untuk mengekstrak fitur (dan juga kata-kata pendapat)



Aturan dari tata bahasa ketergantungan

	Relations and Constraints	Output	Examples
R1 ₁	$O \rightarrow O-Dep \rightarrow F$ s.t. $O \in \{O\}$, $O-Dep \in \{MR\}$, $POS(F) \in \{NN\}$	$f = F$	The phone has a <u>good</u> "screen". $good \rightarrow mod \rightarrow screen$
R1 ₂	$O \rightarrow O-Dep \rightarrow H \leftarrow F-Dep \leftarrow F$ s.t. $O \in \{O\}$, $O/F-Dep \in \{MR\}$, $POS(F) \in \{NN\}$	$f = F$	"iPod" is the <u>best</u> mp3 player. $best \rightarrow mod \rightarrow player \leftarrow subj \leftarrow iPod$
R2 ₁	$O \rightarrow O-Dep \rightarrow F$ s.t. $F \in \{F\}$, $O-Dep \in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	same as R1 ₁ with screen as the known word and good as the extracted word
R2 ₂	$O \rightarrow O-Dep \rightarrow H \leftarrow F-Dep \leftarrow F$ s.t. $F \in \{F\}$, $O/F-Dep \in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	same as R1 ₂ with iPod is the known word and best as the extract word.
R3 ₁	$F_{i(j)} \rightarrow F_{i(j)}-Dep \rightarrow F_{j(i)}$ s.t. $F_{j(i)} \in \{F\}$, $F_{i(j)}-Dep \in \{CONJ\}$, $POS(F_{i(j)}) \in \{NN\}$	$f = F_{i(j)}$	Does the player play dvd with <u>audio</u> and "video"? $video \rightarrow conj \rightarrow audio$
R3 ₂	$F_i \rightarrow F_i-Dep \rightarrow H \leftarrow F_j-Dep \leftarrow F_j$ s.t. $F_i \in \{F\}$, $F_i-Dep = F_j-Dep$, $POS(F_j) \in \{NN\}$	$f = F_j$	Canon "G3" has a great <u>len</u> . $len \rightarrow obj \rightarrow has \leftarrow subj \leftarrow G3$
R4 ₁	$O_{i(j)} \rightarrow O_{i(j)}-Dep \rightarrow O_{j(i)}$ s.t. $O_{j(i)} \in \{O\}$, $O_{i(j)}-Dep \in \{CONJ\}$, $POS(O_{i(j)}) \in \{JJ\}$	$o = O_{i(j)}$	The camera is <u>amazing</u> and "easy" to use. $easy \rightarrow conj \rightarrow amazing$
R4 ₂	$O_i \rightarrow O_i-Dep \rightarrow H \leftarrow O_j-Dep \leftarrow O_j$ s.t. $O_i \in \{O\}$, $O_i-Dep = O_j-Dep$, $POS(O_j) \in \{JJ\}$	$o = O_j$	If you want to buy a <u>sexy</u> , "cool", accessory- available mp3 player, you can choose iPod. $sexy \rightarrow mod \rightarrow player \leftarrow mod \leftarrow cool$



Fitur sinonim grup (Zhai et al., 2010)

- Fitur yang merupakan domain sinonim harus dikelompokkan bersama.
- Banyak teknik yang bisa digunakan untuk mengatasi masalah ini, misalnya,
 - Pemodelan topik, kemiripan distribusi, dll
- Kami mengusulkan metode pembelajaran semi-supervised
 - Z. Zhai, B. Liu, H. Xu dan P. Jia. Mengelompokkan Fitur Produk Menggunakan Pembelajaran Semi-Supervised dengan Soft-Constraints. *COLING-2010*.



Resolusi Coreference (Bing dan Liu 2010)

- Berbeda dengan resolusi coreference tradisional
 - Penting untuk menyelesaikan objek dan fitur
 - Misalnya., “*Saya membeli kamera **Canon S500** kemarin. **Itu** terlihat indah. Aku mengambil beberapa **foto** tadi malam. **Mereka** yang menakjubkan*”.
- Beberapa karakteristik spesifik pendapat dapat dimanfaatkan untuk akurasi yang lebih baik. Lihat
 - X. Ding dan B. Liu, Menyelesaikan Object dan Atribut Coreference dalam Opinion Mining. *Coling-2010*.



Identifikasi orientasi opini

- Untuk setiap fitur, kami mengidentifikasi orientasi sentimen atau opini yang diungkapkan oleh reviewer.
- Hampir semua pendekatan menggunakan kata-kata dan ungkapan opini. Tapi perhatikan lagi (cara yang sederhana):
 - Beberapa kata pendapat memiliki orientasi konteks yang independen, misalnya "hebat".
 - Beberapa kata pendapat lain memiliki orientasi konteks yang bergantung, misalnya, "kecil"
 - Banyak cara menggunakan kata-kata pendapat.
- Metode pembelajaran mesin untuk klasifikasi sentimen pada tingkat kalimat dan klausa juga berlaku.



Agregasi kata-kata pendapat (Bing dan Liu, 2008)

- **Input:** sepasang (f, s) di mana f adalah fitur produk dan s adalah kalimat yang berisi f .
- **Output:** apakah pendapat atas f di s positif, negatif, atau netral.
- Dua langkah:
 - Langkah 1: split kalimat jika diperlukan berdasarkan kata-kata BUT (tapi, kecuali itu, dll).
 - Langkah 2: bekerja pada segmen sf mengandung f . Biarkan set kata-kata pendapat di sf menjadi w_1, \dots, w_n . Singkatnya orientasi mereka $(1, -1, 0)$, dan menetapkan orientasi ke (f, s) sesuai.
- Dalam (Ding et al, WSDM-08), langkah 2 diubah menjadi

$$\sum_{i=1}^n \frac{w_i \cdot o}{d(w_i, f)}$$

dengan hasil yang lebih baik. $w_i \cdot o$ adalah orientasi pendapat w_i . $d(w_i, f)$ adalah jarak dari f to w_i



Aturan Dasar Opini (Liu, Ch dalam buku pegangan NLP)

- Pendapat diatur oleh beberapa peraturan, misalnya,
- Neg | | Negatif
- Pos Positif
- Negasi Neg | | Positif
- Negasi Pos | | Negatif
- Rentang nilai yang diinginkan | | Positif
- Di bawah atau di atas kisaran nilai yang diinginkan | | Negatif



Aturan Dasar Opini (Liu, Ch dalam buku pegangan NLP)

- Penurunan Neg → Positif
- Turun Pos Pos → Negatif
- Peningkatan Neg Negatif
- Peningkatan Pos Positif
- Mengkonsumsi sumber daya → Negatif
- Menghasilkan sumber daya → Positif
- Mengkonsumsi limbah → Positif
- Menghasilkan limbah → Negatif



Dua jenis utama opini

- ◎ **Pendapat Langsung:** Ekspresi sentimen langsung pada beberapa objek target, misalnya produk, acara, topik, orang.
 - Misalnya, "kualitas gambar kamera ini sangat bagus."
 - (banyak yang jauh lebih kompleks).
- ◎ **Pendapat Komparatif:** Perbandingan yang mengungkapkan persamaan atau perbedaan lebih dari satu objek. Biasanya menyatakan pemesanan atau preferensi.
 - Misalnya, "mobil x lebih murah daripada mobil y."



Pendapat Komparatif (Jindal dan Liu, 2006)

⦿ Lemah

- *Tidak-Sama Lemah*: Hubungan yang *lebih besar* jenis atau *kurang dari*
 - Contoh: "optik kamera A lebih baik dari pada kamera B"
- *Equative*: Hubungan dari jenis yang sama dengan
 - Ex: "Sebuah kamera dan kamera B keduanya datang di 7MP"
- *Superlatif*: Hubungan yang *lebih besar* jenis atau *kurang dari semua orang lain*
 - Ex: "Sebuah kamera adalah kamera termurah yang tersedia di pasar"



Pendapat Komparatif Pertambahan

- **Tujuan:** Mengingat dokumen dogmatis d ,. **Ekstrak pendapat komparatif:**

$$(O_1, O_2, F, po, h, t),$$

mana O_1 dan O_2 adalah objek set yang dibandingkan berdasarkan fitur mereka bersama F , po adalah himpunan objek yang disukai dari pendapat pemegang h , dan t adalah waktu ketika pendapat komparatif dinyatakan.

- **Catatan:** bukan pendapat positif atau negatif.



Opini Spam Deteksi (Jindal dan Liu, 2007, 2008)

Ulasan palsu / tidak benar:

Tulisan ulasan positif yang tidak layak untuk beberapa objek target agar bisa mempromosikannya.

Tulisan ulasan negatif yang tidak adil atau berbahaya untuk beberapa objek target untuk merusak reputasi mereka.

Meningkatnya jumlah pelanggan yang mewaspadai ulasan palsu (review bias, ulasan berbayar)



Contoh Praktik Review Spam

Belkin International, Inc

- Produsen jaringan dan periferal teratas | Penjualan ~ \$ 500 juta di tahun 2008
- Memposkan iklan untuk menulis ulasan palsu di amazon.com (65 sen per ulasan)

Timer: 00:00:00 of 60 minutes

Want to work on this HIT? Want to see other HITs?

Write Product Reviews 25-50 Words
Requester: Mike Bayard
Qualifications Required: HIT approval rate (%) is not less than 95

Jan 2009

Write a Positive 5/5 Review for Product on Website

Positive review writing.

- Use your best possible grammar and write in US English only
- Always give a 100% rating (as high as possible)
- Keep your entry between 25 and 50 words
- Write as if you own the product and are using it
- Tell a story of why you bought it and how you are using it
- Thank the website for making you such a great deal
- Mark any other negative reviews as "not helpful" once you post yours

Instructions:

The link below leads to a product on a website. Read-through the product's features and write a positive review for it using the guidelines above to the best of your ability. I have also provided the part number for this product and you can click on the links below to see it on several alternative websites. In order to post some reviews you will need to create an account on the site. You can use your own email address or open a new free webmail account (gmail, yahoo...) and use it to post with.



Percobaan dengan Ulasan Amazon

● Juni 2006

- Ulasan 5.8mil, 1.2mil produk dan 2.1mil pengulas.

● Ulasan memiliki 8 bagian

- *<ID Produk> <ID Reviewer> <Rating> <Date> <Review Title> <Review Body> <Jumlah Umpan Balik yang Bermanfaat> <Jumlah Umpan Balik> <Jumlah Umpan Bermanfaat>*

● Industri diproduksi produk “mProducts”

● **misalnya elektronik, komputer, aksesoris, dll**

- Ulasan 228K, 36K produk dan ulasan 165K.



Beberapa Hasil Tentatif

- Ulasan outlier negatif cenderung sangat spammed.
- Ulasan tersebut merupakan satu-satunya ulasan beberapa produk yang kemungkinan akan diblokir
- Peninjau peringkat teratas lebih cenderung menjadi spammer.
- Ulasan spam bisa mendapatkan umpan balik bermanfaat yang bagus dan ulasan non-spam bisa mendapatkan masukan buruk.



Pertemuan Ilmu Sosial

- Ekstrak dan analisis pendapat politik.
 - Kandidat dan isu
- Bandingkan pendapat di seluruh budaya dan lang.
 - Membandingkan pendapat orang dari berbagai negara mengenai isu atau topik yang sama, misalnya diplomasi internet
- Opini spam (pendapat palsu)
 - Apa itu sosial, budaya, aspek ekonomi itu?
- Propagasi opini dalam konteks sosial
- Bagaimana pendapat di Web mempengaruhi dunia nyata
 - Apakah mereka berkorelasi?
- Analisis emosi dalam konteks sosial & dunia maya



References

- B. Liu, “Sentiment Analysis and Subjectivity.” A Chapter in *Handbook of Natural Language Processing*, 2nd Edition, 2010.
 - (An earlier version) B. Liu, “Opinion Mining”, A Chapter in the book: *Web Data Mining*, Springer, 2006.



Kesimpulan & Review

- ◎ Kami secara singkat mendefinisikan masalah analisis sentimen.
 - Pendapat langsung: fokus pada analisis tingkat fitur
 - Pendapat komparatif: berbagai jenis perbandingan
 - Opini deteksi spam: ulasan palsu
 - Saat ini bekerja dengan Google (Google research award).
- ◎ Banyak aplikasi.
- ◎ Tantangan teknis masih besar.
 - Tapi kami cukup optimis.
- ◎ Tertarik untuk berkolaborasi dengan ilmuwan sosial
 - pendapat dan isu terkait secara inheren bersifat sosial.



Kuis (Latihan Soal)

- Carilah paper tentang Sistem Temu Kembali Informasi (STKI) minimal 3 paper yang berkaitan dengan ***Opinion Mining (Sentiment Analysis)***.
- Paper wajib menjelaskan mengenai tahapan-tahapan yang berkaitan.
- Rangkumlah menjadi artikel baru tentang review masalah diatas (***minimal 2 halaman kertas A4***).



Thanks!

Any questions ?