

Election Monitor 2004

Information Retrieval - Semantic Technologies

Usecase

Albert Weichselbraun

- US Election 2004
- Monitoring via Websites
- Indikatoren: Keywords, Semantic Orientation, Attention
- weitere Anwendung: Image von Produkten/Firmen/..., political Monitoring, ...

INFORMATION RETRIEVAL

- 1 -

28. NOVEMBER 2007

George W. Bush
Republican (born July 6, 1946)

This project of the **ECORESEARCH.NET** automatically provides a weekly snapshot of international web coverage. The results reflect online attention and attitude towards the US presidential candidates. Keywords grouped by political party and geographic region summarize the key issues associated with each candidate. [more]

| Region | United States | Canada | United Kingdom | Australia & NZ |
|---------------|---------------|---------|----------------|----------------|
| Attention (%) | 66 3 | 68 -3 | 74 0 | 70 -2 |
| Attitude | 2 0 | 1 1 | 0 0 | 1 0 |

Top 10 Keywords

| | | | |
|--------------|------------|------------|----------------|
| speech | iraq | iraq | blair |
| iraq | war | victory | iraq |
| gore | visit | tony | war |
| war | crimes | stave | tony |
| blair | clinton | blair | white |
| term | americans | war | administration |
| secretary | pennsylv | america | runsfeld |
| washington | washington | washington | speech |
| inauguration | fahrenheit | washington | ronald |
| white | reagan | speech | ohio |

Attention and Attitude of US Media Sites

| Candidates | Attention (%) | Attitude | Top 10 Keywords |
|---------------|---------------|----------|-----------------|
| Bush, George | 71 3 | 1 -1 | speech |
| Camejo, Peter | 26 -3 | -3 1 | iraq |
| Cheney, Dick | | | rally |
| Edwards, John | | | gore |
| Kerry, John | | | debate |
| Nader, Ralph | | | secretary |

Top 10 Keywords

| | | | |
|------------|---------------|---------------|------------|
| speech | nominee | vietnam | debate |
| iraq | ballot | percent | holds |
| rally | trading | liberty | speaks |
| gore | debate | secretary | challenger |
| debate | secretary | massachusetts | rally |
| war | massachusetts | ohio | concession |
| washington | ohio | concession | supporters |
| blair | concession | supporters | interviews |
| term | supporters | interviews | |

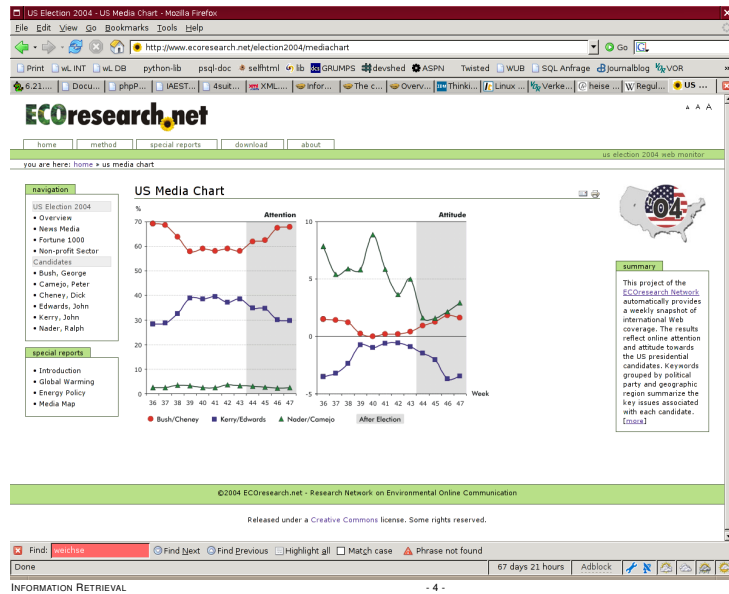
INFORMATION RETRIEVAL

- 3 -

28. NOVEMBER 2007

- 2 -

28. NOVEMBER 2007



Attention

• Basistechnologie: Regular Expressions

- Muster, die auf bestimmten Text passen
- Definition der Muster (alternation, grouping, quantification)
Beispiel: `((Ö|Oe)sterreichs?|Austrias?)`

Regular Expressions

| Zeichen | Bedeutung |
|----------------|---|
| . | passt auf jedes beliebige Zeichen |
| \w | alphanumerische Zeichen (\W) |
| \s | Whitespaces (\S) |
| \d | Zahlen (\D) |
| [a-k] | alle Buchstaben von a-k |
| [a-k0-3] | alle Buchstaben von a-k und Zahlen zwischen 0 und 3 |
| (first second) | Unterscheidung mehrerer Alternativen |

Tabelle 1: Definition von Mustern

Zeichen Bedeutung

| | |
|--------|---|
| ^ | Beginn einer Zeile |
| \$ | Ende einer Zeile |
| {m,n} | zwischen m und n Vorkommnisse |
| {m,n}? | zwischen m und n Vorkommnisse, non-greedy |
| * | {0,x} Vorkommnisse |
| + | {1,x} Vorkommnisse |
| ? | {0,1} Vorkommnisse |
| (?...) | nicht gruppierende Version von Klammern |
| (?!..) | nicht |

Tabelle 2: Definition von Mustern

Regular Expressions - Beispiele

```
elections?(-| )?2004
```

```
elections?(-| )?\d{2,4}
```

```
george (?:w |w\. |walker )?bush
```

```
(?:president|pres\.|republican|candidate)
```

```
john (?:f |f\. |forbes )?kerry
```

```
(?:senator|sen\.|democrat|candidate) kerry
```

```
(?!\d+) [\\"\\(\)\.a-z0-9_]{2,}
```

Semantic Orientation

- Wie Medien über Dinge schreiben (positiv/negativ)

- automatisiert berechnet

- Grammatik: Negation

- $SO = \sum \frac{SO_{w_i}}{(d_{w_i})^\lambda}$

- λ abhängig von der Modellierung

- reverse Lemmatisation

Semantic Orientation - Beispiele

- “US stocks *rallied* Wednesday, *boosted* by *shares* of health and *defence companies* that are seen *benefiting* from the re-election of President **George W. Bush**, but higher oil prices checked *advances*”
(NEW ZEALAND HERALD). (+4.09)
- “The dollar *hit* its *lowest* level in more than eight months against the Euro Thursday, *falling* sharply on *worries* about the economic effects of rising oil prices and expectations of continued trade and budget *deficits* in **President Bush**’s second term”
(ST. PETERSBURG TIMES). (-4.03)

Keyword Analysis

- Vergleich: Referenzcorpus/Zielcorpus

→ Document-Counts

→ χ^2 -Test; $\chi_{n-1}^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$

→ Yates Korrektur für Kontinuität: $\chi_{Yates}^2 = \sum_{i=1}^n \frac{(|O_i - E_i| - 0.5)^2}{E_i}$;

- ermittelt Wörter/Konzepte, die statistisch signifikant öfters mit dem Basiskonzept gemeinsam vorkommen

- Pre-Processing:

- Text in Sätze/Dokumente splitten

- Tokenization

- Word Stemming (optional)