OpenGeist
Insight in the Stream of Page Views on Wikipedia

Maria-Hendrike Peetz, Edgar Meij, Maarten de Rijke
An answer to (one of) Fernando’s question
Original data versus Trend

ratio of blog posts labeled with STRESSED

06/03/05 08/01/05 10/01/05 12/01/05 02/01/06 04/01/06
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
IR researcher @ A university

- Semantic analysis of tweets
  - “What are they about?”

- Semantic analysis of live tv
  - Provide background information

- Reputation management project
  - Reliable statistics
  - Normalization
OpenGeist

- Understanding the zeitgeist
- Zeitgeist? What is that?
  - http://www.en.wikipedia.org/zeitgeist
- OpenGeist is a RESTful API to access wikipedia page views and provide basic analysis
What’s in OpenGeist

- **Raw**
- **Statistics**
- **Grouping**
  - Of related pages/concepts
  - Of categories
- **Counts, plots**

http://opengeist.org/{begindate}/{enddate}
*{resolution}/*{filter}*
[/category/{id}]
[/raw]
[/topN[/compare/
*{filter_graph}]]
[/cluster[/k/
*{filter_graph}]]
[/statistics]
[/concept/{id}]
[/raw]
[/topN]
[/cluster[/k/
*{filter_graph}]]
[/compare/
*{filter_graph}]
[/statistics]
Uses so far

- “Semantification”
  - Query logs
  - Tweets
  - TV feeds
OpenGeist

- Public API
  - http://www.opengeist.org/about
  - http://www.opengeist.org/examples

- (Some firewall issues today)