Exploratory Search Missions for TREC Topics

EuroHCIR 2013

Dublin, 1 August 2013

Martin Potthast  Matthias Hagen  Michael Völske  Benno Stein

Bauhaus-Universität Weimar

www.webis.de
Background

- Dataset for studying text reuse / plagiarism from web sources
- PAN workshop at CLEF
- Previous datasets: real plagiarism or algorithmically generated
- This time: crowdsourcing to freelance authors
Background

Corpus Construction

Editor

ChatNoir SE

ClueWeb API

Revision log

Query log

ClueWeb

Barack Obama's Family

The Family of Barack Obama is an extended clan of African American, Indonesian, and Kenyan heritage. They are best known through the political career of Barack Obama, the current President of the United States. His immediate family is the First Family of the United States, the first First Family of African American descent in the United States' history, and since the Kennedy era, Obama's family has been back in the news because of his family ties to the Kennedy family.

In what follows, we give a detailed overview of Barack Obama's family, his immediate and extended family, including matrilineal relations. Moreover, we give insights into the relations of Michelle Obama, a well-known woman, as well as some distant relations of both.

Barack Obama

Barack Hussein Obama II is the 44th and current President of the United States, the first African American to hold the office. Obama was a junior senator from Illinois from 2005 until he resigned following his election.
The Family of Barack Obama is an extended clan of African American, Indonesian, and Kenyan heritage. They are best known through the political career of Barack Obama, the current President of the United States. His immediate family is the First Family of the United States, the first First Family of African American descent in the United States to enter the White House since the Kennedys. Obama's extended family harks back to the days of Camelot.

In what follows, we give a detailed overview of Barack Obama's family background. His immediate and extended family, including their relations, are provided. Moreover, we give insights into the relations of Michelle Obama's wife, as well as some distant relations of both.

Barack Obama
Barack Hussein Obama II is the 44th and current President of the United States, the first African American to hold the office. Obama was the junior Senator from Illinois from 2005 until he resigned following his election.
Example topic:

*Obama’s family.*

Write about President Barack Obama’s family history, including genealogy, national origins, places and dates of birth, etc. Where did Barack Obama’s parents and grandparents come from? Also include a brief biography of Obama’s mother.

Original topic 001 of the TREC Web Track 2009:

*Query.* obama family tree

*Description.* Find information on President Barack Obama’s family history, including genealogy, national origins, places and dates of birth, etc.

*Sub-topic 1.* Find the TIME magazine photo essay “Barack Obama’s Family Tree.”

*Sub-topic 2.* Where did Barack Obama’s parents and grandparents come from?

*Sub-topic 3.* Find biographical information on Barack Obama’s mother.
Background

Authors

Revision log

Query log

ClueWeb

ChatNoir SE

ClueWeb API

Editor

The Family of Barack Obama is an extended clan of African American, Indonesian, and Kenyan heritage. They are best known through the political career of Barack Obama, the current President of the United America. His immediate family is the First Family of the United States, the first First Family of African American descent in the United States. The youngest to enter the White House since the Kennedys, Obama's family harks back today's Celltool.

In what follows, we give a detailed overview of Barack Obama's Farnington on himself, his immediate and extended family, including maten relations. Moreover, we give insights into the relations of Michelle O Obama's wife, as well as some distant relations of both.

Barack Obama

Barack Hussein Obama II is the 44th and current President of the United States and the first African American to hold the office. Obama was the junior Senator from Illinois from 2005 until he resigned following his election.
Background

Authors

- 12 Professional writers hired on oDesk [www.odesk.com]
- Fluent English speakers with several years writing experience

Author Demographics

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Median)</td>
<td>37</td>
</tr>
<tr>
<td>Years Writing (Median)</td>
<td>8</td>
</tr>
<tr>
<td>Academic degree</td>
<td></td>
</tr>
<tr>
<td>Postgrad</td>
<td>33%</td>
</tr>
<tr>
<td>Undergrad</td>
<td>25%</td>
</tr>
<tr>
<td>English</td>
<td></td>
</tr>
<tr>
<td>Native</td>
<td>67%</td>
</tr>
<tr>
<td>Second Language</td>
<td>33%</td>
</tr>
</tbody>
</table>
Background

Search engine and query log

Editor

ChatNoir SE

Revision log

Query log

ClueWeb API

ClueWeb
Background

Clueweb search engine

The Family of Barack Obama in Kenya
http://www.aztlan.net/barack_obama_family.htm
Los Angeles, Alta California July 25, 2008LA VOZ DE AZTLAN The Family of Barack Obama in Kenya The father of Barack Obama was married to Kezia Obama in Kenya prior to marrying the senator’s mother Ann

Barack Obama and Family on Election Night
http://www.gazoube.com/tn/MO7/RvAC.html
- Mickey v - TubeFire.com - 3GP FM Barack Obama Sheds Tears for His Beloved Grandmother 0:37 Behind The Scene Picel Barack Obama Family Election Night 0:41 Raw Images Obama on Election Night 1:30 Obama and family snorkel at Hanauma Bay 3:03 Barack Oba

Bargain Prices On First Families
http://www.retrostats.com/stats/First-Families.aspx
1st Steps for Family Fren $ 10.95 The First Family Obama pin! 3" NEW LOWER PRICES!!!! $ 0.99 *OFFICIAL* THE FIRST FAMILY BARACK OBAMA 3" BUTTON PIN $ 4.99 Second City: First Family of Comedy - New DVD LOW PRICE $ 15.55 Pokemon the First Movie-Mewtwo V

Bargain Prices On First Families
ER PRICES!!!! $ 0.99 *OFFICIAL* THE FIRST FAMILY BARACK OBAMA 3" BUTTON PIN $ 4.99
Second City: First Fa

Yes We Can - Blog
http://www.yeswecanweibtbands.us/blog/tag/obama-family
o the current White House pootch, Barney Bush! Woof, Woof! jason | Post a Comment | Share
Article tagged obama dog, obama family barack obama white house, dog in U.S. POLITICS This
Background
Clueweb search engine

- Source retrieval for essay writing
- Indexes ClueWeb09
- Fine-grained interaction logs
## Query logs

### Basic stats

<table>
<thead>
<tr>
<th>Corpus Characteristic</th>
<th>Distribution</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min</td>
<td>avg</td>
</tr>
<tr>
<td>Queries</td>
<td>13</td>
<td>651</td>
</tr>
<tr>
<td>Queries / Topic</td>
<td>4</td>
<td>91.0</td>
</tr>
<tr>
<td>Clicks</td>
<td>16</td>
<td>739</td>
</tr>
<tr>
<td>Clicks / Topic</td>
<td>12</td>
<td>111.6</td>
</tr>
<tr>
<td>Sessions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sessions / Topic</td>
<td>1</td>
<td>12.3</td>
</tr>
<tr>
<td>Days</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days / Topic</td>
<td>1</td>
<td>4.9</td>
</tr>
<tr>
<td>Hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours / Writer</td>
<td>3</td>
<td>129.3</td>
</tr>
<tr>
<td>Hours / Topic</td>
<td>3</td>
<td>7.5</td>
</tr>
</tbody>
</table>
Findings so far
Findings so far
Distribution of Queries Over Time

<table>
<thead>
<tr>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>64</td>
<td>42</td>
<td>64</td>
<td>8</td>
<td>36</td>
</tr>
<tr>
<td>B</td>
<td>76</td>
<td>102</td>
<td>12</td>
<td>66</td>
<td>80</td>
</tr>
<tr>
<td>C</td>
<td>99</td>
<td>48</td>
<td>274</td>
<td>69</td>
<td>111</td>
</tr>
<tr>
<td>D</td>
<td>198</td>
<td>94</td>
<td>88</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>E</td>
<td>57</td>
<td>34</td>
<td>70</td>
<td>36</td>
<td>136</td>
</tr>
<tr>
<td>F</td>
<td>319</td>
<td>153</td>
<td>148</td>
<td>248</td>
<td>109</td>
</tr>
</tbody>
</table>

Distribution of queries over time.

- fraction of posed queries (y-axis) over elapsed time (x-axis) between the first query until essay completion
- each cell represents one of 150 essays
- the numbers denote the total amount of posed queries
- the cells are sorted by area under the curve
### Findings so far

**Distribution of Queries Over Time**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Distribution of queries over time.

- fraction of posed queries (y-axis) over elapsed time (x-axis) between the first query until essay completion
- each cell represents one of 150 essays
- the numbers denote the total amount of posed queries
- the cells are sorted by area under the curve
Findings so far
Distribution of Queries Over Time

- Distribution of queries over time.
  - fraction of posed queries (y-axis) over elapsed time (x-axis) between the first query until essay completion
  - each cell represents one of 150 essays
  - the numbers denote the total amount of posed queries
  - the cells are sorted by area under the curve
Summary

1. Corpus of essay writing, focus on text reuse
2. 150 exploratory search missions for TREC web track topics
3. First appraisal indicates wide spectrum of search behavior
4. Data will be publicly available: http://webis.de/research/corpora
Summary

1. Corpus of essay writing, focus on text reuse
2. 150 exploratory search missions for TREC web track topics
3. First appraisal indicates wide spectrum of search behavior
4. Data will be publicly available: http://webis.de/research/corpora

Thank you for your attention!