

DEAR SEARCH ENGINE: WHAT'S YOUR OPINION ABOUT...?

Sentiment Analysis for Semantic Enrichment of Web Search Results

Gianluca Demartini and Stefan Siersdorfer

L3S Research Center

April 2010

MOTIVATION

- As the Web grows, Search Engines are the main entry points
- The “visible” Web is: top search engine results

- Controversial topics are being discussed on the Web (news,blogs,...)

- Search Engines can bias the way we see the Web!
- It is important to provide a good overview in top-N results
 - Topic
 - Sentiment

OUR CONTRIBUTION

- Approaches for computing sentiment of web pages
- Comparing sentiments for:
 - three search engines
 - top results vs lower ranked

OUTLINE

- Enhancing Search Results with Sentiments
- Related Work
- Extracting Sentiment from Web Pages
- Experimental Study
- Future Work

"There is no quality of life when the patient is dead."

We are committed to the fundamental belief that the intentional killing of another person is wrong.

euthanasia

Search

[Euthanasia suicide mercy-killing right-to-die physician ...](#)

Comprehensive information for research on **euthanasia**, assisted suicide, living wills, and mercy killing.

www.euthanasia.com - [Cached](#)

[Euthanasia suicide mercy-killing right-to-die physician ...](#)

Map of the Places in the World Where **Euthanasia**/ Assisted Suicide is Legal (2008)
Status of **Euthanasia** and Assisted Suicide World Wide (2005) ...

www.euthanasia.com/page4.html - [Cached](#)

[Euthanasia - Wikipedia, the free encyclopedia](#)

[Classification of euthanasia](#) | [Procedural decision](#) | [References](#) | [See also](#)

Euthanasia + θάνατος, thanatos (death)) refers to the practice of ending a life in a painless manner. According to the House of Lords Select Committee on Medical Ethics, the precise definition of **euthanasia** is "a deliberate...

en.wikipedia.org/wiki/Euthanasia - [Cached](#)

[euthanasia: Definition from Answers.com](#)

euthanasia n. The act or practice of ending the life of an individual suffering from a terminal illness or an incurable condition, as by lethal

www.answers.com/topic/euthanasia - 149k - [Cached](#)

[Voluntary Euthanasia \(Stanford Encyclopedia of Philosophy\)](#)

Voluntary Euthanasia. First published Thu Apr 18, 1996; substantive revision Wed Aug 27, 2008. The entry sets out five individually necessary conditions ...

plato.stanford.edu/entries/euthanasia-voluntary

[Euthanasia](#)

Euthanasia: **Euthanasia** is the intentional killing by act or omission ... **Euthanasia** By Action: Intentionally causing a person's death by performing an action such ...

www.nrlc.org/euthanasia/index.html - [Cached](#)

THE IDEAL RESULT

Search engines might rank results according to sentiment

What is the “ideal ranking”?

Possible ideal results might be:

- Balanced Overview (both +1 and -1 docs)
- Neutral Overview (0 docs)
- Realistic Overview (80% negative docs)
- Personalized Overview (as the user likes)

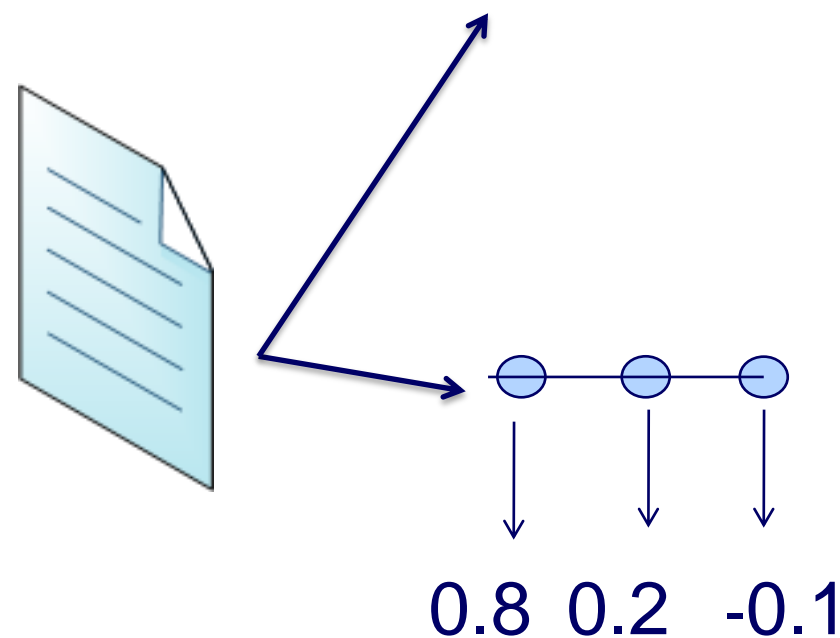
RELATED WORK

- Bias in search engines
 - topical bias
 - on content/url level
- Search results diversification
 - relevance/diversity trade-off
- Opinion Mining
 - mainly on blogs (TREC)
 - approaches use Machine Learning techniques
- Enriching Search Results

EXTRACTING SENTIMENT FROM WEB PAGES

Sentiment Classification

- Lexicon-based: SentiWordNet
- Text Classification: Linear SVM trained with movie reviews



$$pos(st) = \frac{\sum_{a \in Adjectives(st)} pos(a)}{|Adjectives(st)|}$$

$$neg(st) = \frac{\sum_{a \in Adjectives(st)} neg(a)}{|Adjectives(st)|}$$

$$obj(st) = \frac{\sum_{a \in Adjectives(st)} obj(a)}{|Adjectives(st)|}$$

SENTIMENT ESTIMATION METHODS

Aggregation

- Given a web page
 - Full text
 - Relevant Sentences
- Assign to it
 - Class with highest score (2 or 3 classes)
 - Real classification score

Methods:

- | | |
|---------------|---------------|
| ■ SWN_SEN_BIN | ■ SVM_SEN_BIN |
| ■ SWN_SEN_MAX | ■ SVM_SEN_REA |
| ■ SWN_TXT_BIN | ■ SVM_TXT_BIN |
| ■ SWN_TXT_MAX | ■ SVM_TXT_REA |

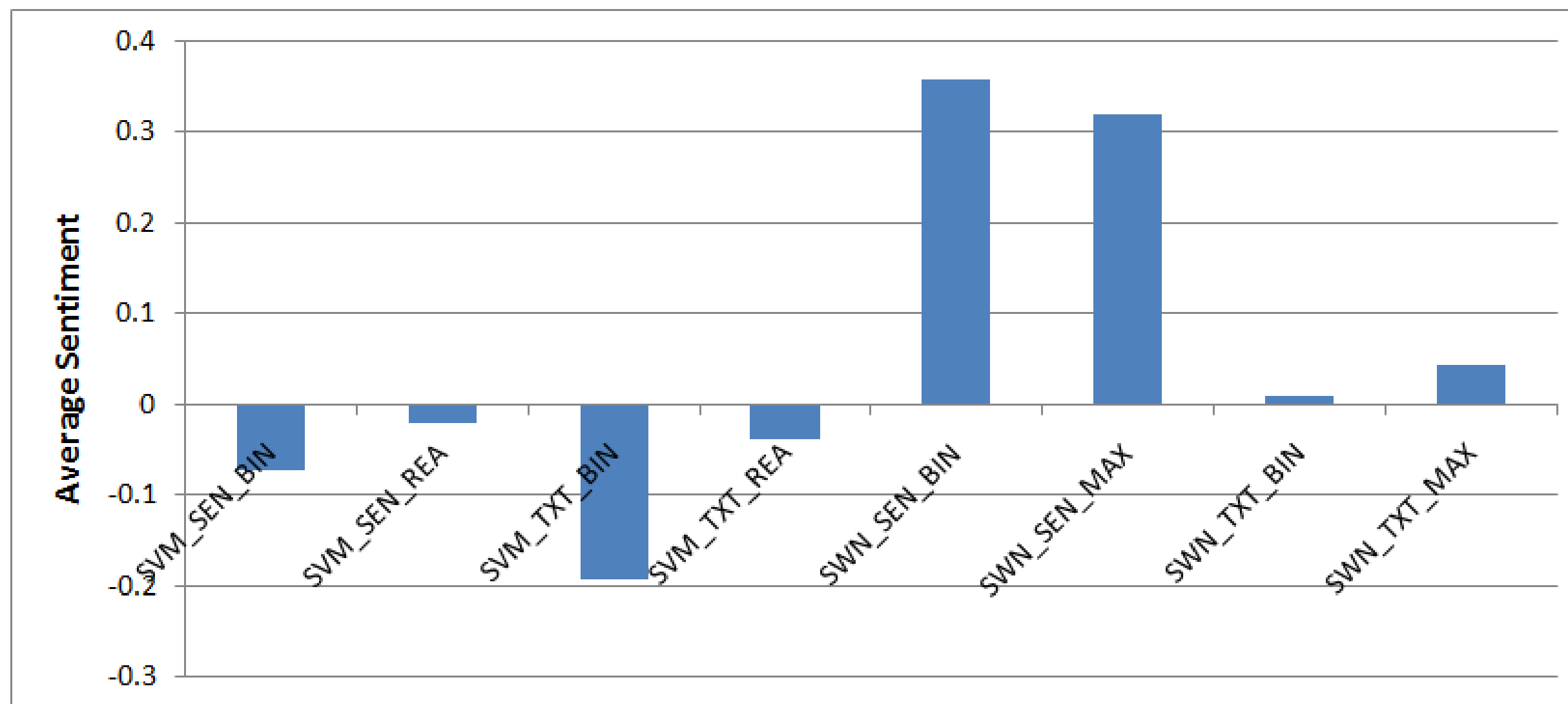
EXPERIMENTAL ANALYSIS

- 3 search engines: Bing, Google, Yahoo!
- 14 opinionated queries
- 50 search results

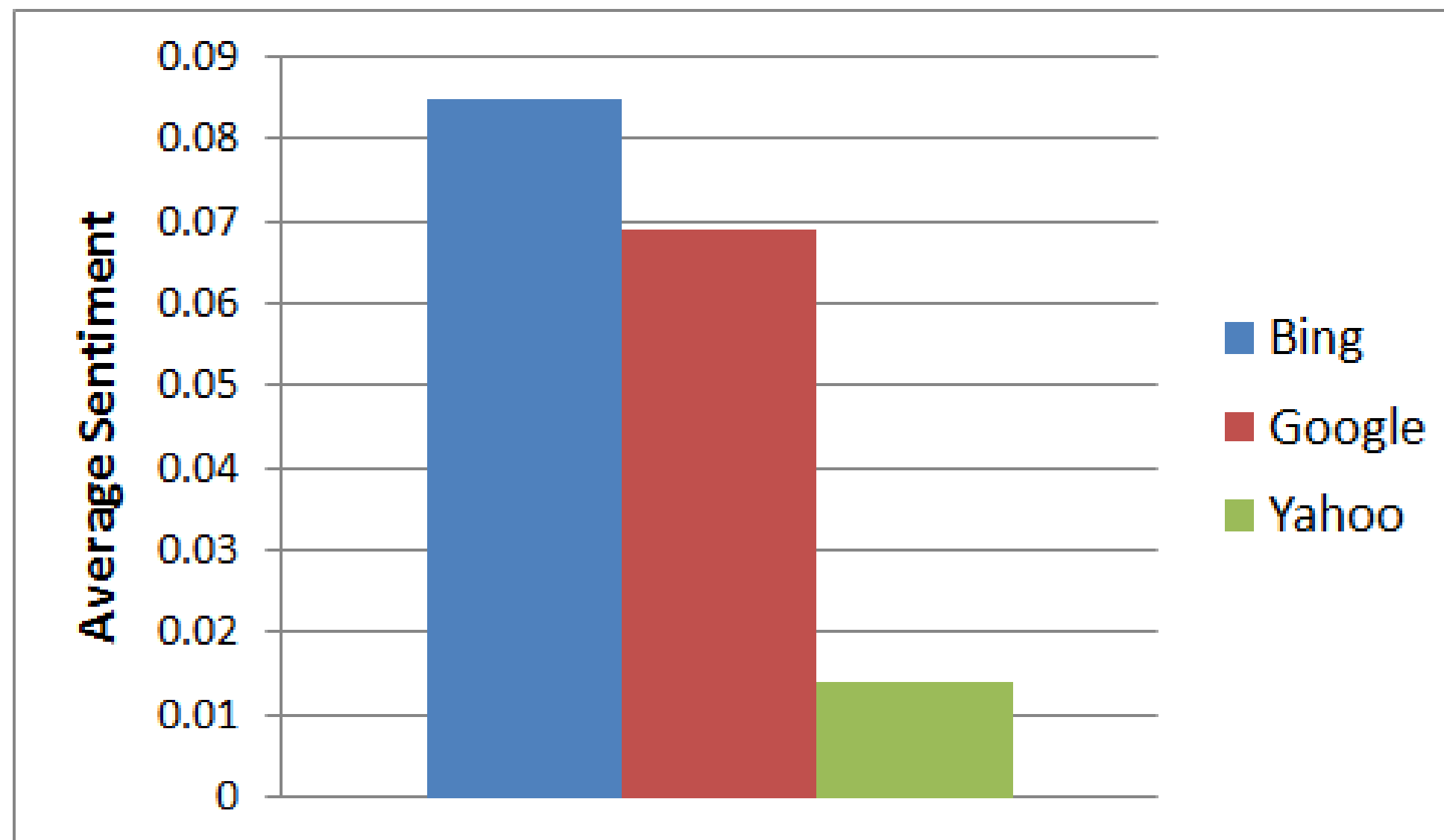
abortion	immigration
anorexia	islam
cloning	marijuana
economy	marriage
employment	nazism
euthanasia	vegetarianism
homosexuality	vividown

METHOD COMPARISON

SEN more positive than TXT

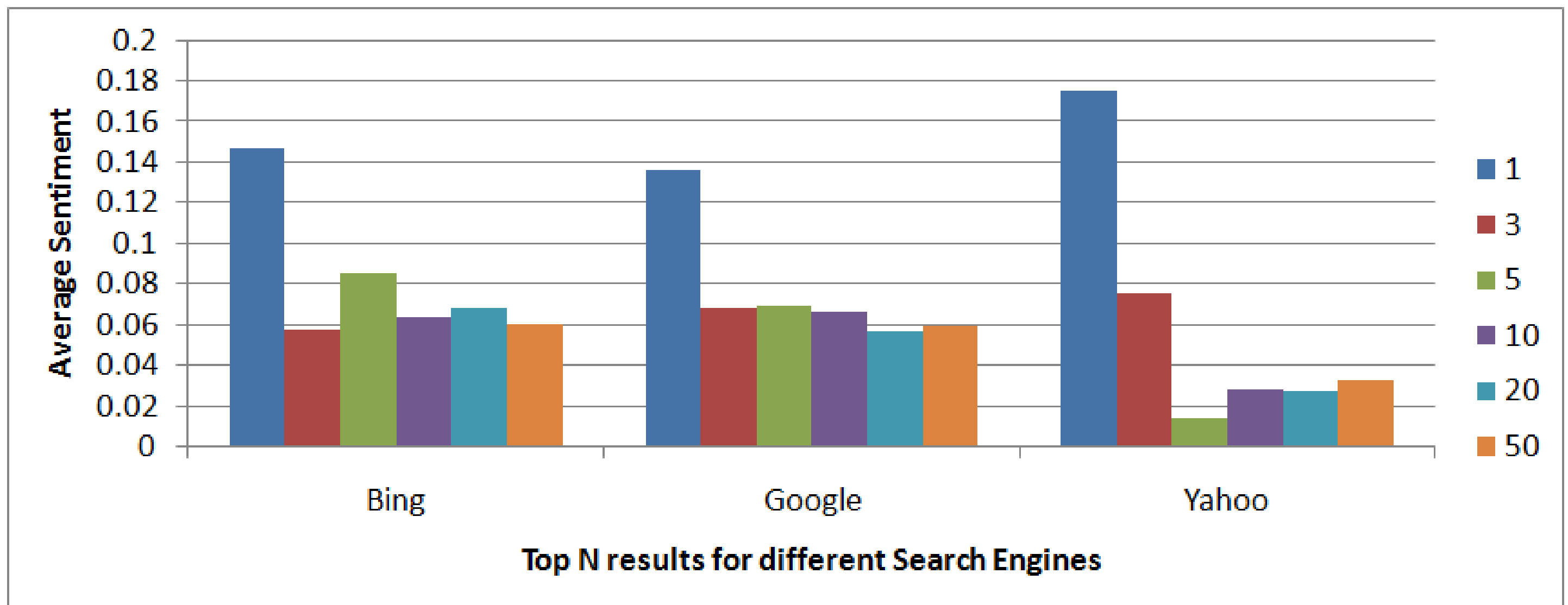


SENTIMENT FOR DIFFERENT SE



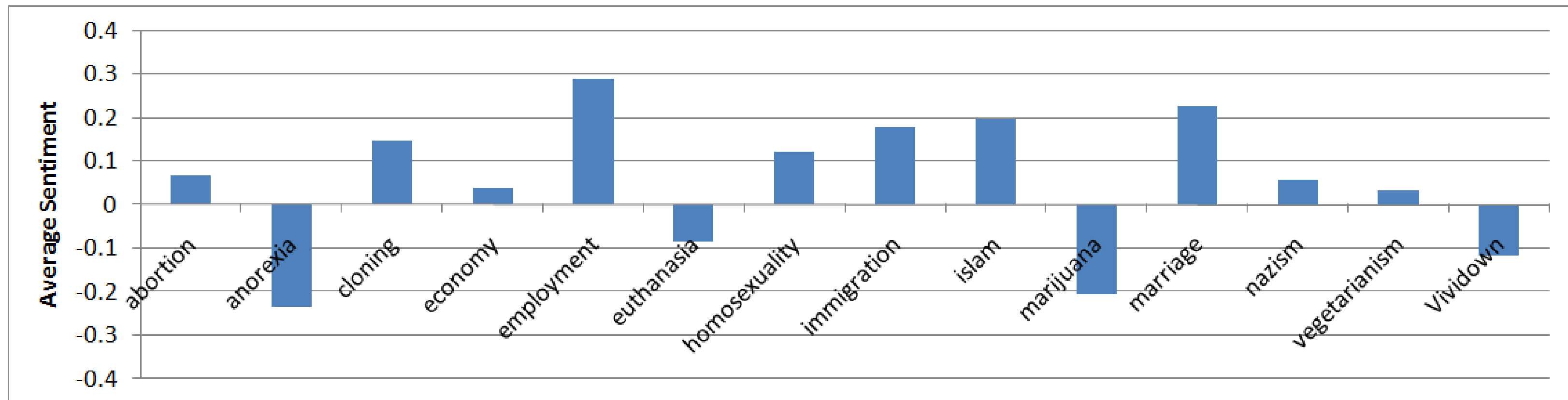
SENTIMENT AT DIFFERENT RANKS

Top-1 is usually the “home page”



SENTIMENT ABOUT DIFFERENT TOPICS

Senti(Employment) > Senti(Marijuana)



FUTURE WORK

- Metadata Annotations: Explain what is the opinion in top results
- Measure effectiveness of different sentiment classification methods
- User Study to judge the usefulness of sentiment diversification
- Sentiment Diversification
 - Detect if top N results are a good sample of other results
 - Evaluate Search Engines
 - How topic/sentiment diversity interact

FUTURE WORK

Applications:

- Re-ranking techniques for producing a fair top-n result set (NP-hard)
- Fairness of Wikipedia (should be objective!)
- Lab experiments: fairness of BM25 on TREC ClueWeb corpus
- Search Results VS Public Opinion (Search engine VS blogosphere)

