

---

# Is Tagging Effective? – Overlapping Ratios with Other Metadata Fields

---

Wooseob Jeong

School of Information Studies

University of Wisconsin – Milwaukee

---

# Introduction

- The potential advantages of tagging have been addressed in numerous venues including monographs and popular magazines as well as academic journals.
- A much smaller number of studies have questioned the effectiveness of tagging, motivation of tagging, and the dominant “personal” as opposed to the “social” aspect of tagging, along with the reaffirmation of the value of traditional controlled vocabulary.
- However, few pointed out that significant overlapping exists between tagging and other existing metadata fields, such as title and description, which makes tagging lose its effectiveness.

---

# Purpose of Study

- Unlike the majority of current research on tagging, which supports tagging's potential, this study questions the effectiveness of it, because of the redundancy with additional findings from the data.

---

# Research Questions

- Do the numbers of unique words in metadata increase over time?
- Are there any differences between videos with web site promotion and those without such promotion with regards to the numbers of unique words in metadata?
- Are there any changes over time in the ratio of overlapping between unique words among metadata fields?

---

# Methodology (1)

- Videos from Youtube.com
- Filtering
  - One video per member
  - Videos with only one or no URL included in their description field to avoid spam
  - A total of 17,130 videos' metadata were extracted initially.
- Word counting
  - “Unique” words in the fields of title, description and tag were counted.
  - Varied forms of a word were counted as separate words: ex) “books” and “book”
  - Special characters were removed and replaced by a space before word counting: ex) “father’s” became “father” and “s.”
- Outliers ( $SD > 2.5$ ) were excluded: 16,084 videos remained in the data set.

---

# Methodology (2)

- 8 ratios were calculated: the percentage of words
  - RTD: from the description in the title
  - RDT: from the title used in the description
  - RDG: from the tags used in the description
  - RGD: from the description used in the tag
  - RTG: from the tags used in the title
  - RGT: from the title used in the tag
  - RCG: from the tag used in the title-description combination
  - RGC: from the title-description combination used in the tag

# Example of Data

The image shows a screenshot of a YouTube video page in Internet Explorer. The browser's address bar shows the URL <http://youtube.com/watch?v=yVPlixS2HMU>. The page title is "YouTube - Barack Obama in Milwaukee, WI". The YouTube logo and navigation tabs (Home, Videos, Channels, Community) are visible at the top. A search bar and an "Upload" button are also present.

The video player shows a man in a suit speaking at a podium in front of a crowd. The video title, "Barack Obama in Milwaukee, WI", is circled in red. The video player controls show a progress bar at 00:25 / 02:26.

Below the video player, the video's metadata is displayed:

- Rate: ★★★★★ (340 ratings)
- Views: 44,288
- Comments: 62
- Favorited: 79 times
- Honors: 0
- Links: 5

On the right side of the page, there is a channel banner for "OBAMA'08" with the website [www.barackobama.com](http://www.barackobama.com). The channel information includes:

- From: [BarackObamadotcom](http://BarackObamadotcom)
- Joined: 1 year ago
- Videos: 802
- Subscribe button

The "About This Video" section contains the following text:

Barack spoke to a crowd of 6,000 supporters in Milwaukee on February 15, 2008. ([less](#))

Added: February 15, 2008

Category: News & Politics

Tag: [barack](#) [obama](#) [milwaukee](#) [wisconsin](#)

The URL is <http://youtube.com/watch?v=yVPlixS2HMU>.

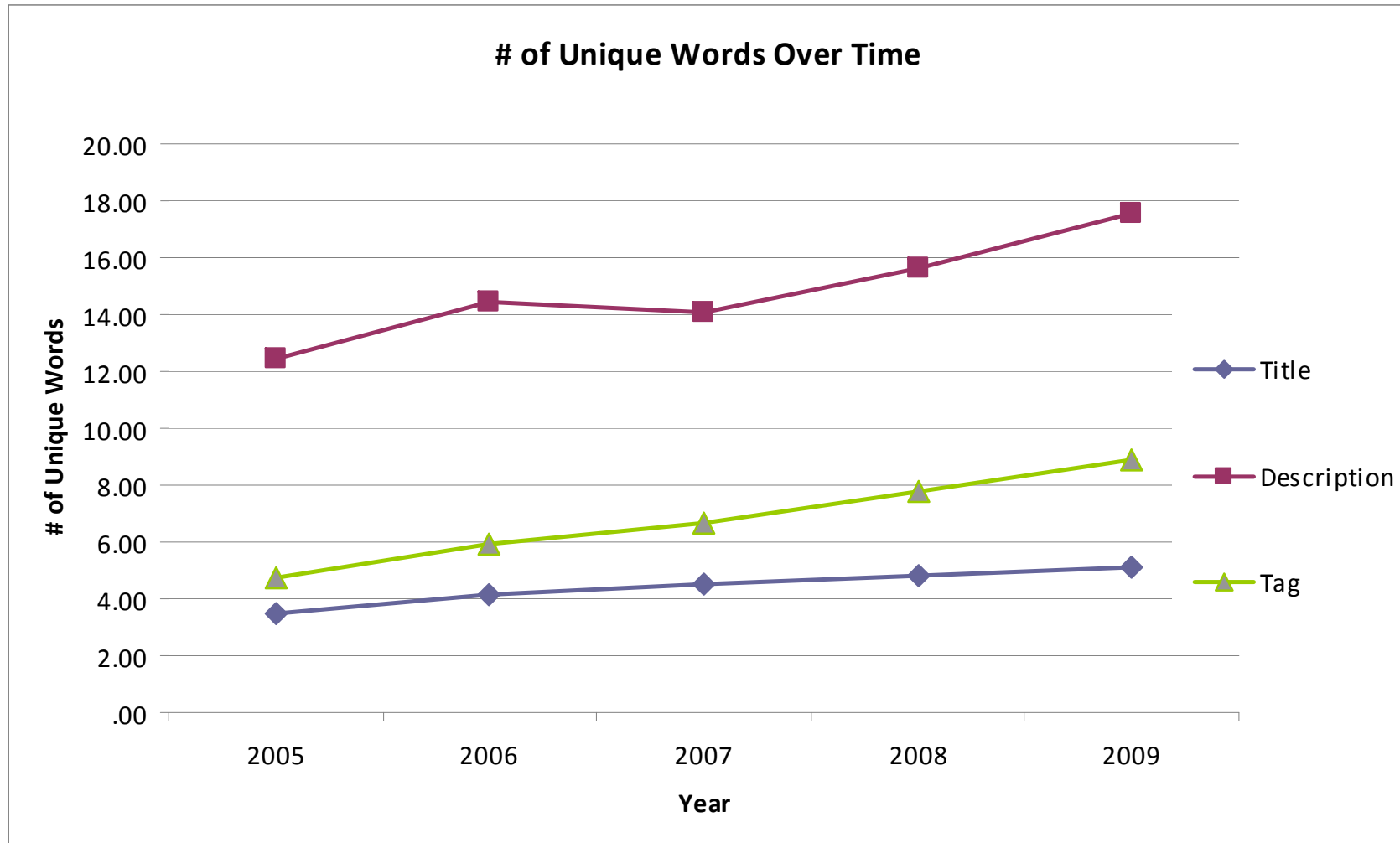
There is an "Embed" section with a "customize" link and an embed code: `<object width="425" height="355"><param name="movie" value="http://`

Below the embed code is a "Make a contribution to this candidate" section with a "Choose Amount" dropdown and a "Contribute" button with a "Google Checkout" logo.

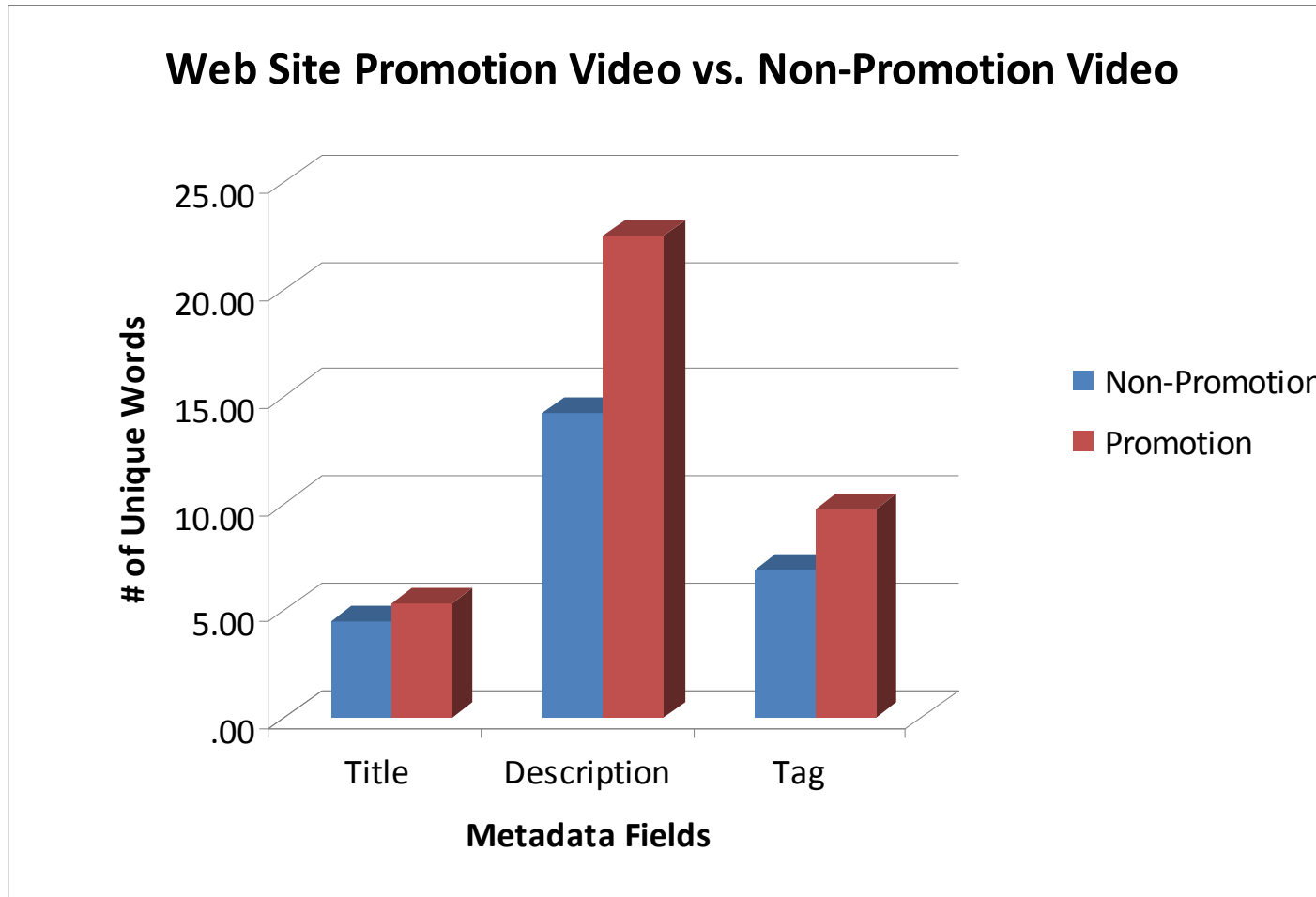
At the bottom of the page, there is a "More From: BarackObamadotcom" section and a "QuickList" section with 1 video and a "Play All" link.

The Windows taskbar at the bottom shows the system tray with the time 9:38 PM and the system clock.

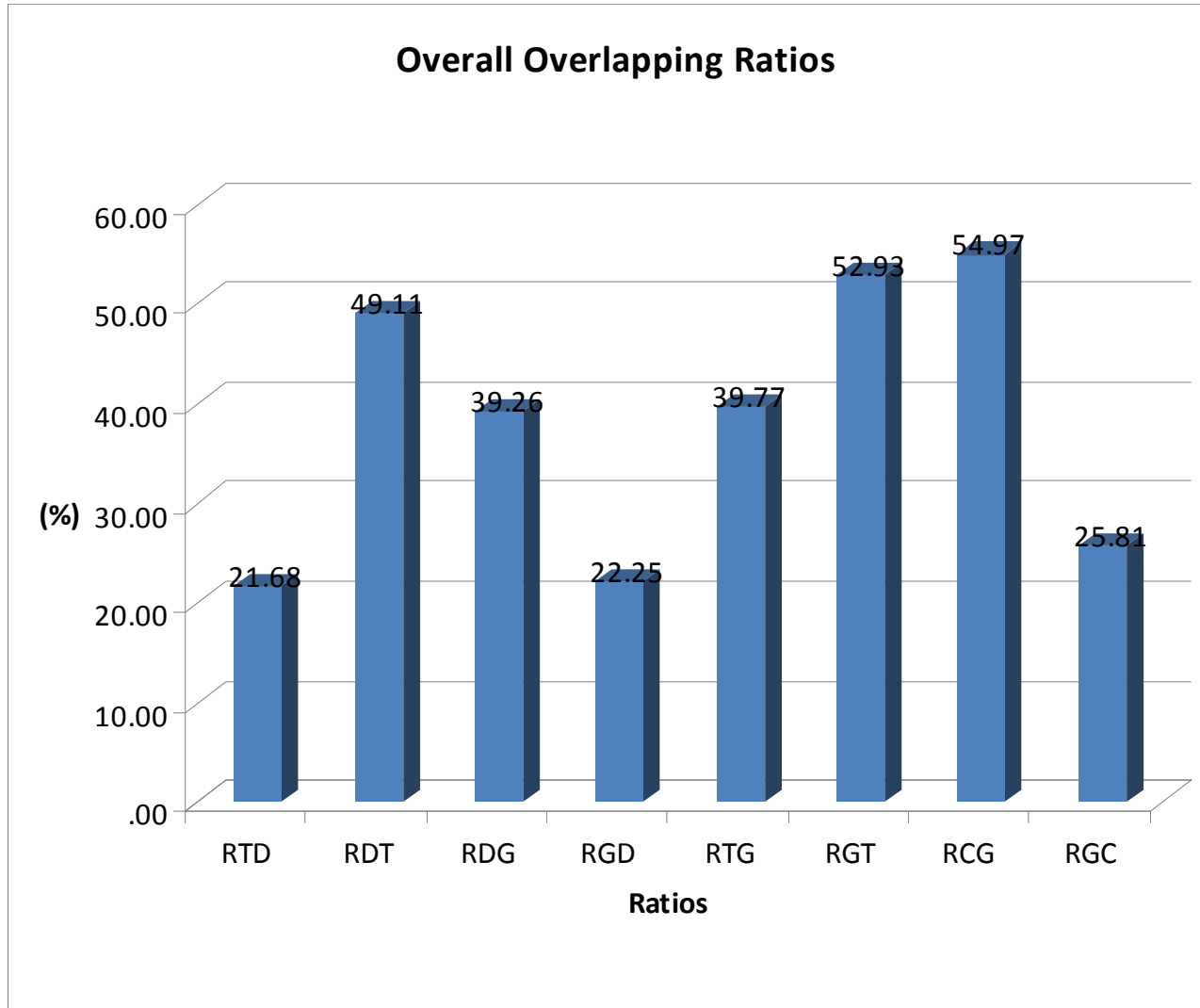
# Numbers in Metadata Fields Over Time



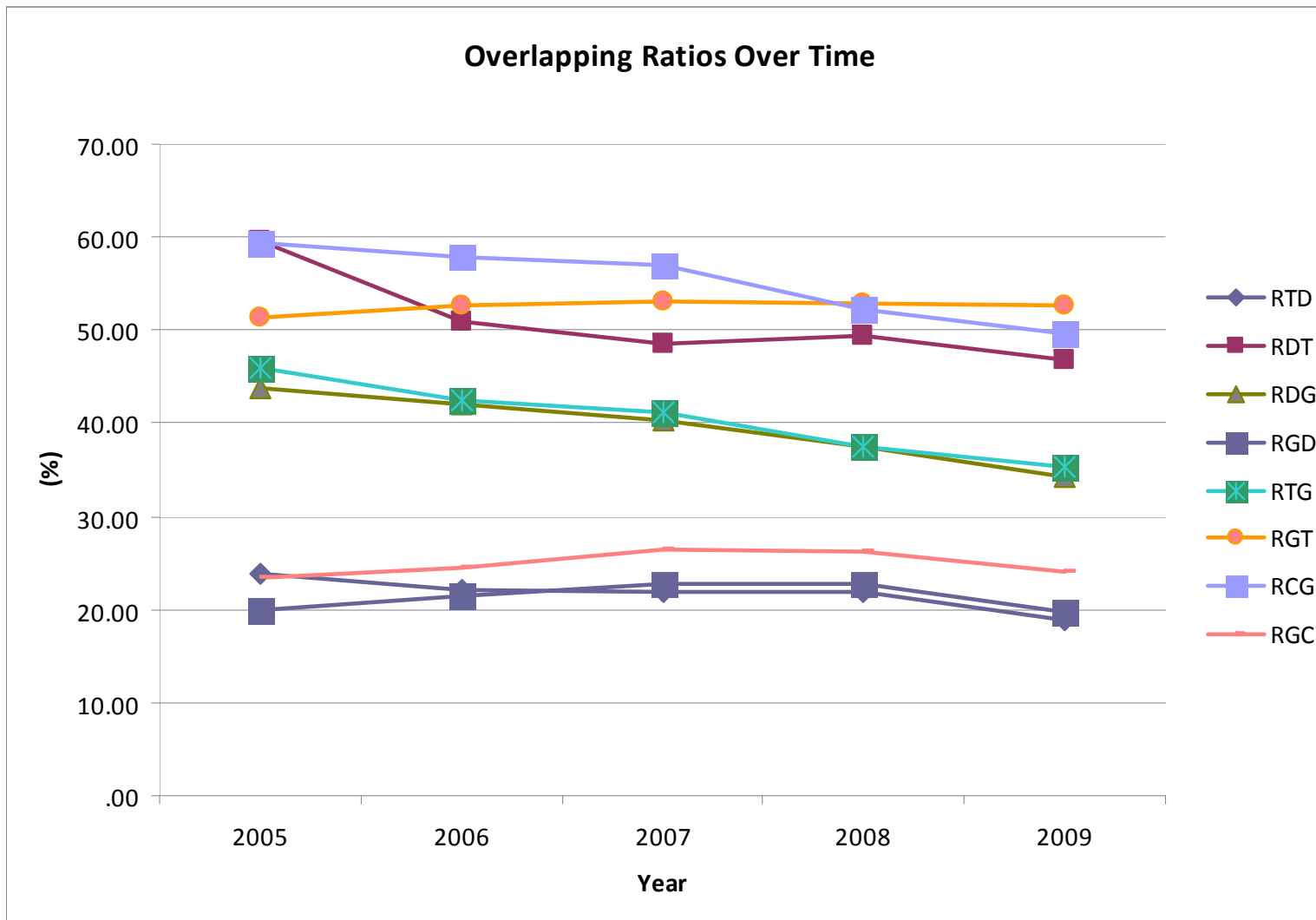
# Difference between Videos Promoting a Web Site and Those with No Web Site



# Overlapping Ratios among Metadata



# Overlapping ratios over time



---

# Absolute Overlapping Ratios

- As much as 25% of the videos have the exactly same words repeated among the metadata fields, such as between the description and the title, and between the tag and the title.

---

# Discussion (1)

- Significant overlapping ratios were found among the fields of title, description and tag for videos at Youtube.com.
- With more aggressive word counting, such as collapsing plurals and tenses into a single word, the overlapping percentage could be higher.

---

## Discussion (2)

- The “new” concept of creators tagging their work does not seem to show much improvement, due to its significant overlapping with the existing metadata fields like title and description.
- Much richer information in both fields would help the search performance without the overlapped words in the tag field.

---

## Discussion (3)

- The technical aspect of tagging is rather disappointing, especially for not allowing multi-word tags.
- Requiring multiple fields of metadata can be a burden to authors, submitters, and indexers.
- Other observations during the data collection and analysis

---

# Conclusion

- This study shows that significant overlapping ratios exist among the metadata fields of title, description, and tag in videos of Youtube.com, questioning the effectiveness of tagging for organization and retrieval of information.
- It also shows additional pattern of tagging, such as increased numbers of words used for each field, difference between web site promotion videos and other videos, and the changes in overlapping ratios among fields over time.

---

# Acknowledgements

- Many thanks should be directed to Soohyung Joo, Hohyon Ryu, Jaehyun Park, Jihee Beak, and Jeanette C Jordan, for their assistance in the data collection and the data analysis for this paper.