

# How Writers Search

Analyzing the Search and Writing Logs of Non-fictional Essays

CHIIR 2016

16 March 2016

---

Martin Potthast      Matthias Hagen      **Michael Völske**      Jakob Gomoll      Benno Stein  
Bauhaus-Universität Weimar

[www.webis.de](http://www.webis.de)

<firstname>.<lastname>@uni-weimar.de

# How Writers Search

Analyzing the Search and Writing Logs of Non-fictional Essays

CHIIR 2016

16 March 2016

---

- Outline
- Our Dataset: Webis-TRC-12
  - Searching and Writing Strategies
  - Query Formulation

# Webis-TRC-12 (Text Reuse Corpus)

- Text reuse from web sources; lack of public data.
- Source retrieval as well as paraphrasing.
- Beyond text reuse.
- Essay writing task with content reuse; recording search engine interactions as well as text progress.

# Dataset Construction

# Dataset Construction

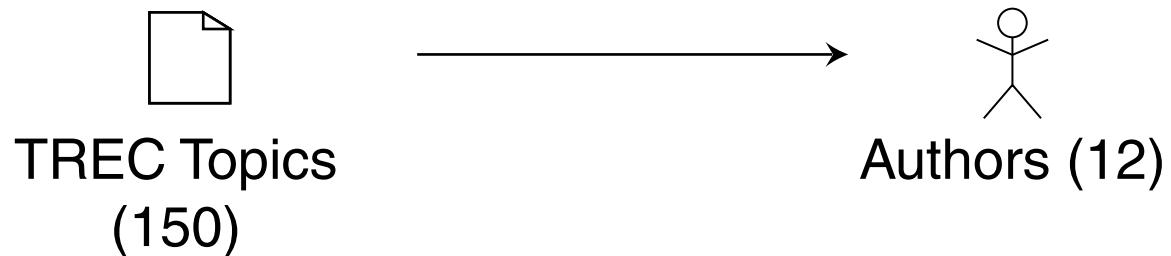


TREC Topics  
(150)

ClueWeb09



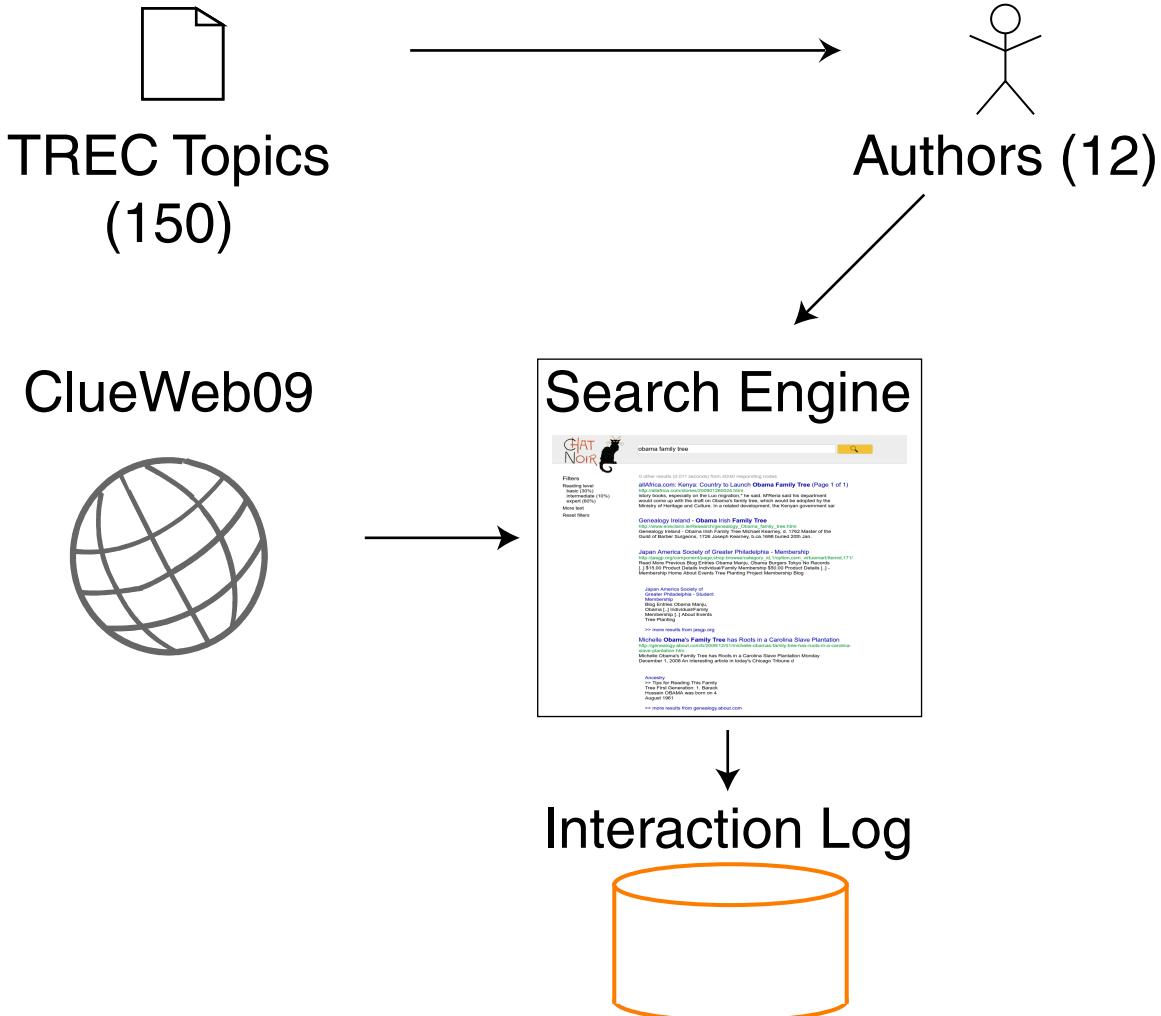
# Dataset Construction



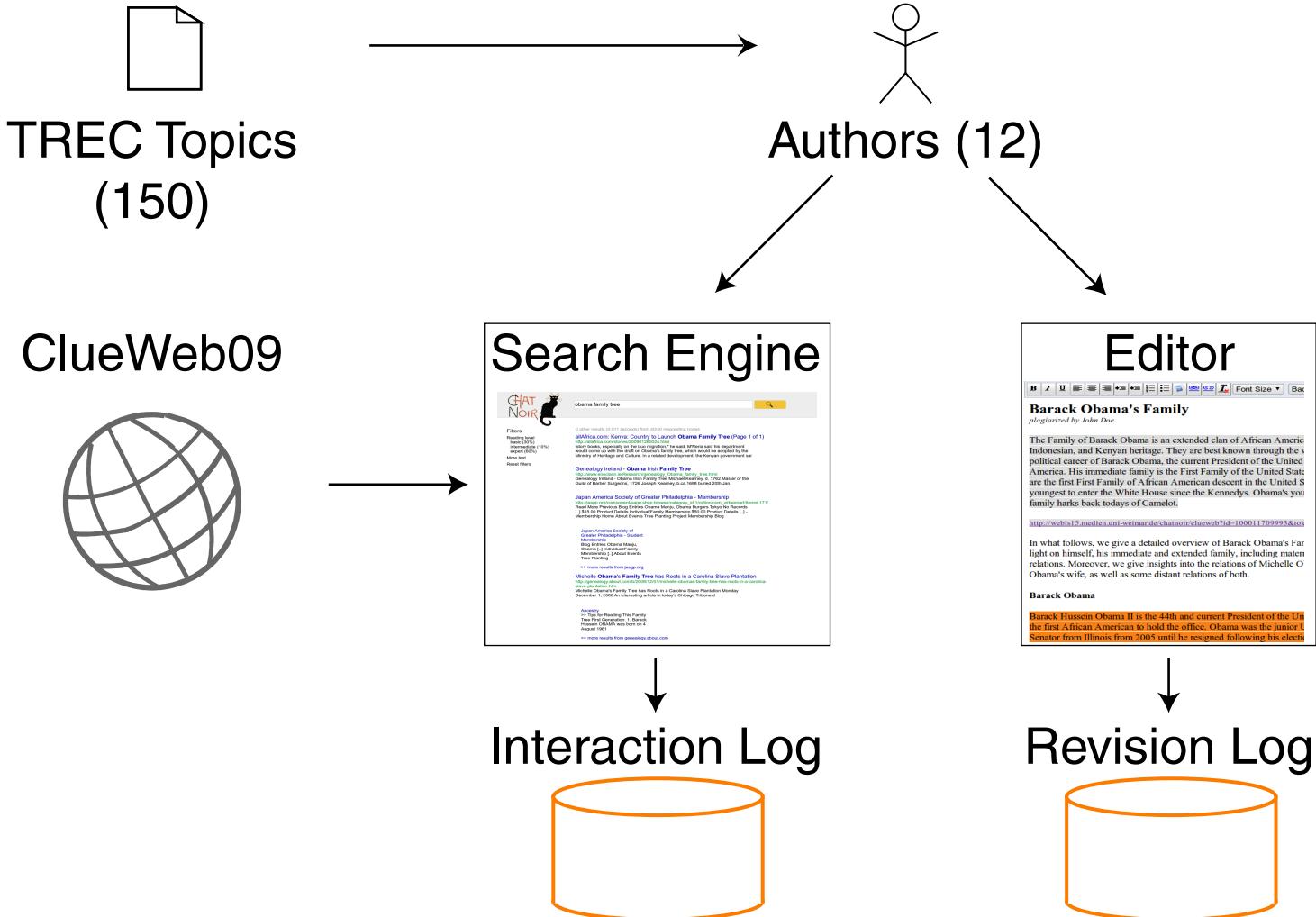
ClueWeb09



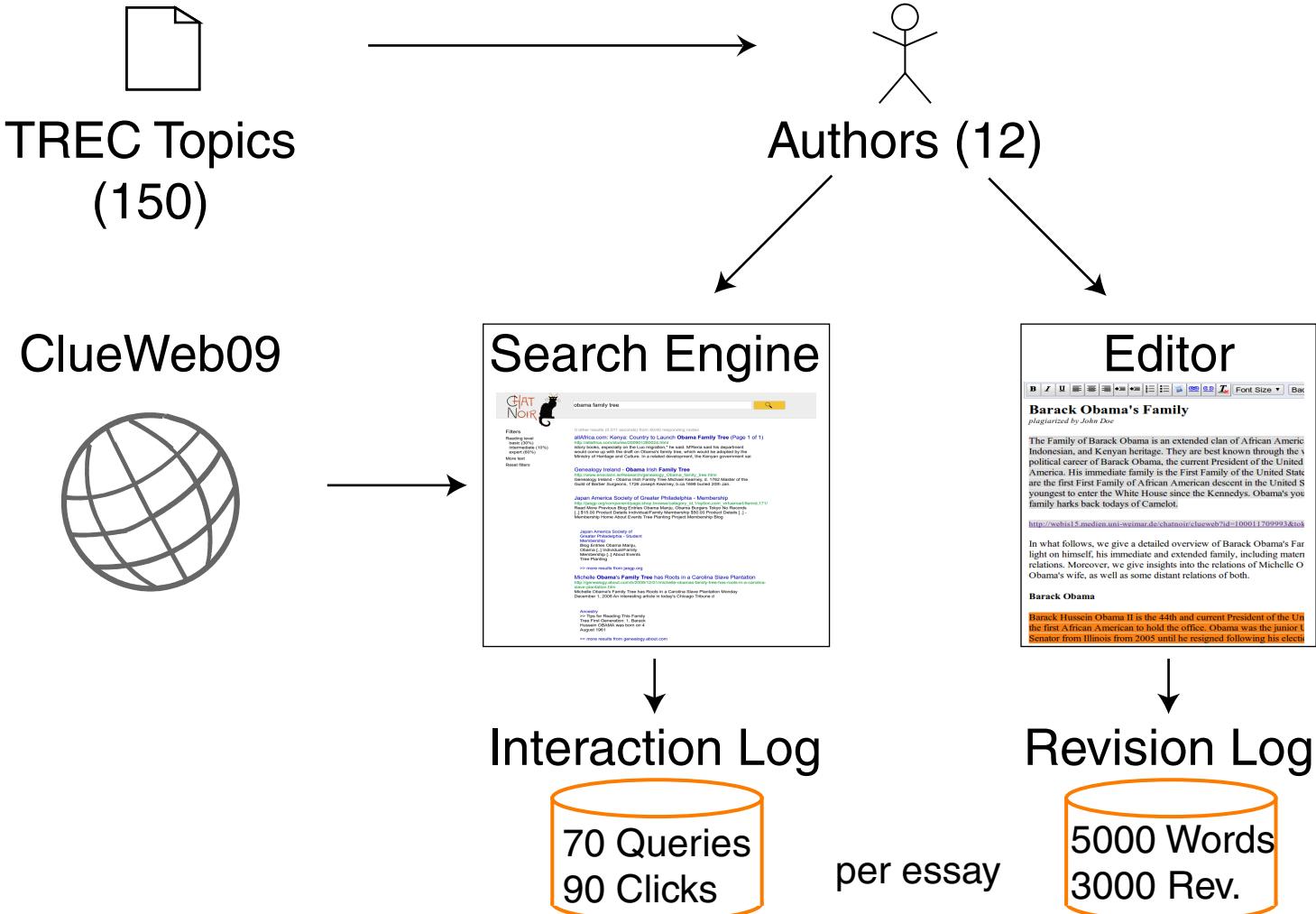
# Dataset Construction



# Dataset Construction



# Dataset Construction

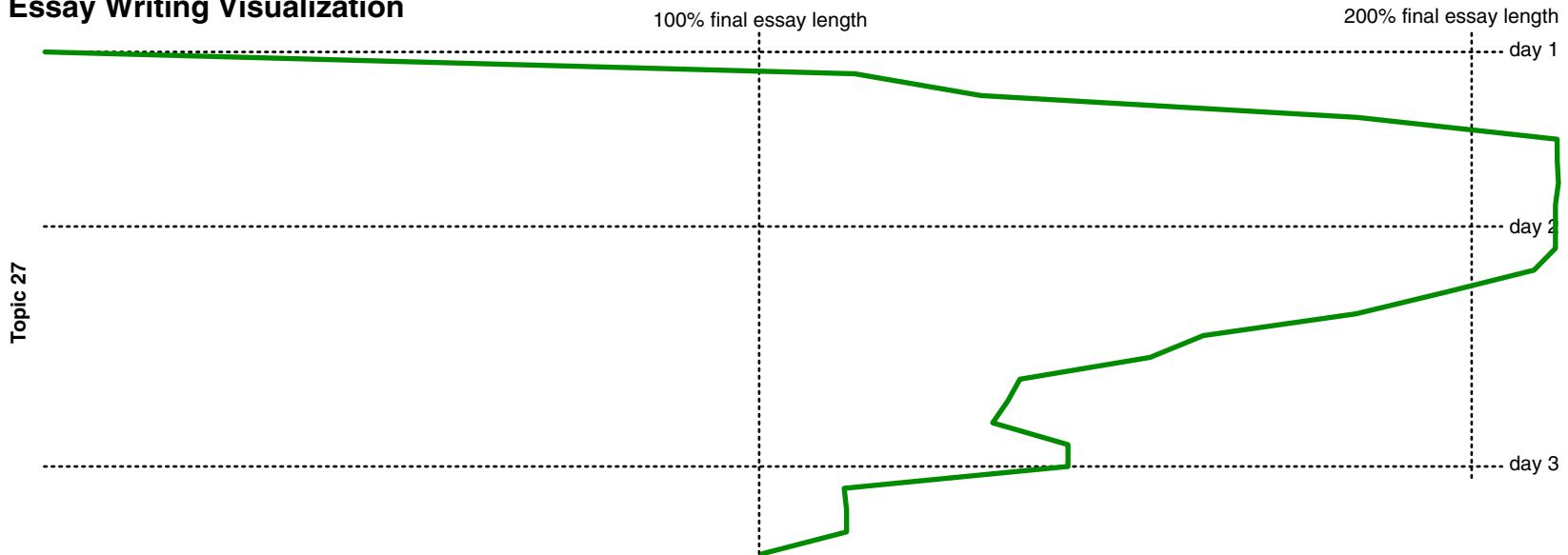


# Writing and Search Behavior

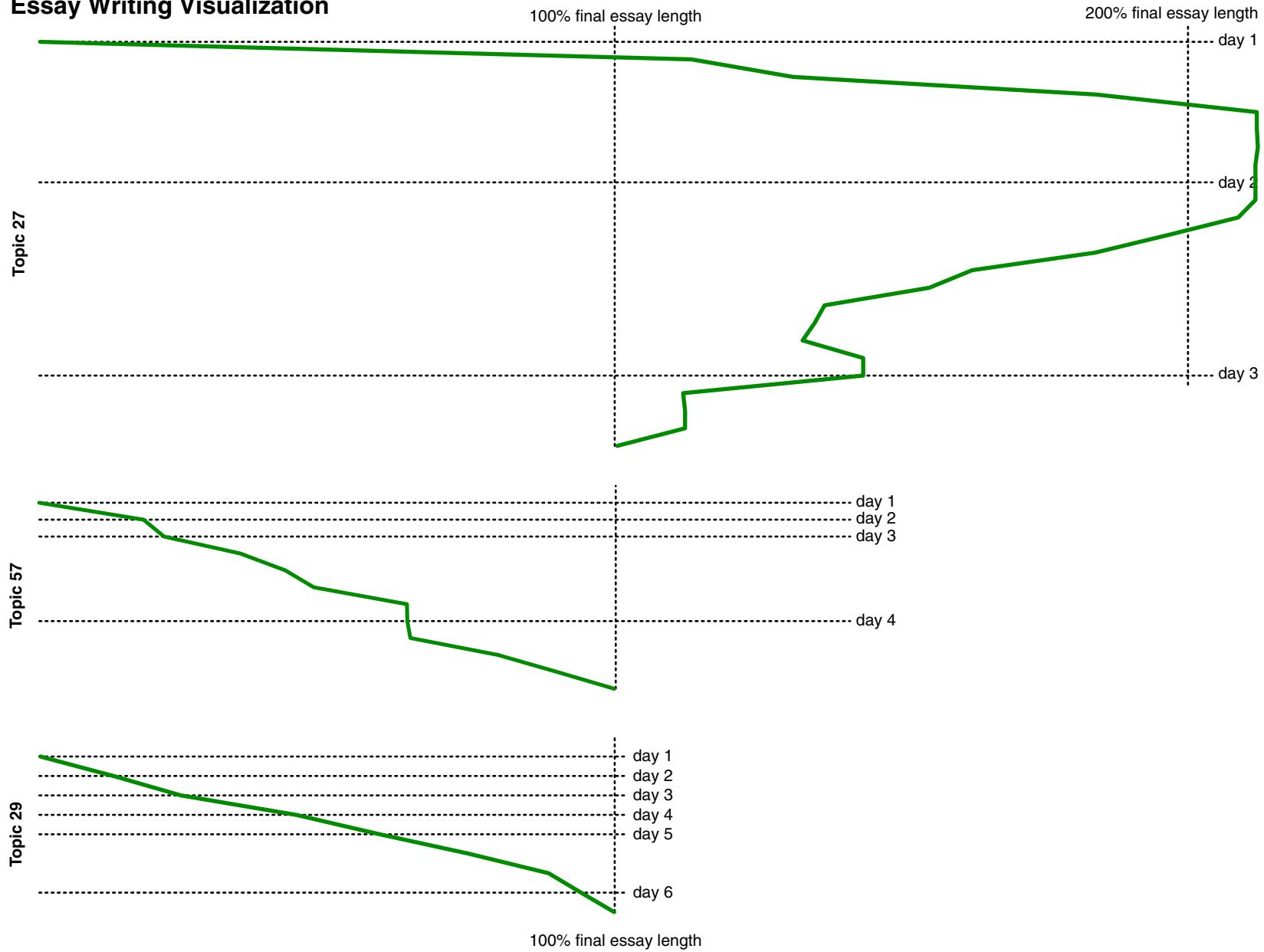
## Research Questions

1. Are different writing and searching strategies distinguishable?
2. How consistent are the authors?

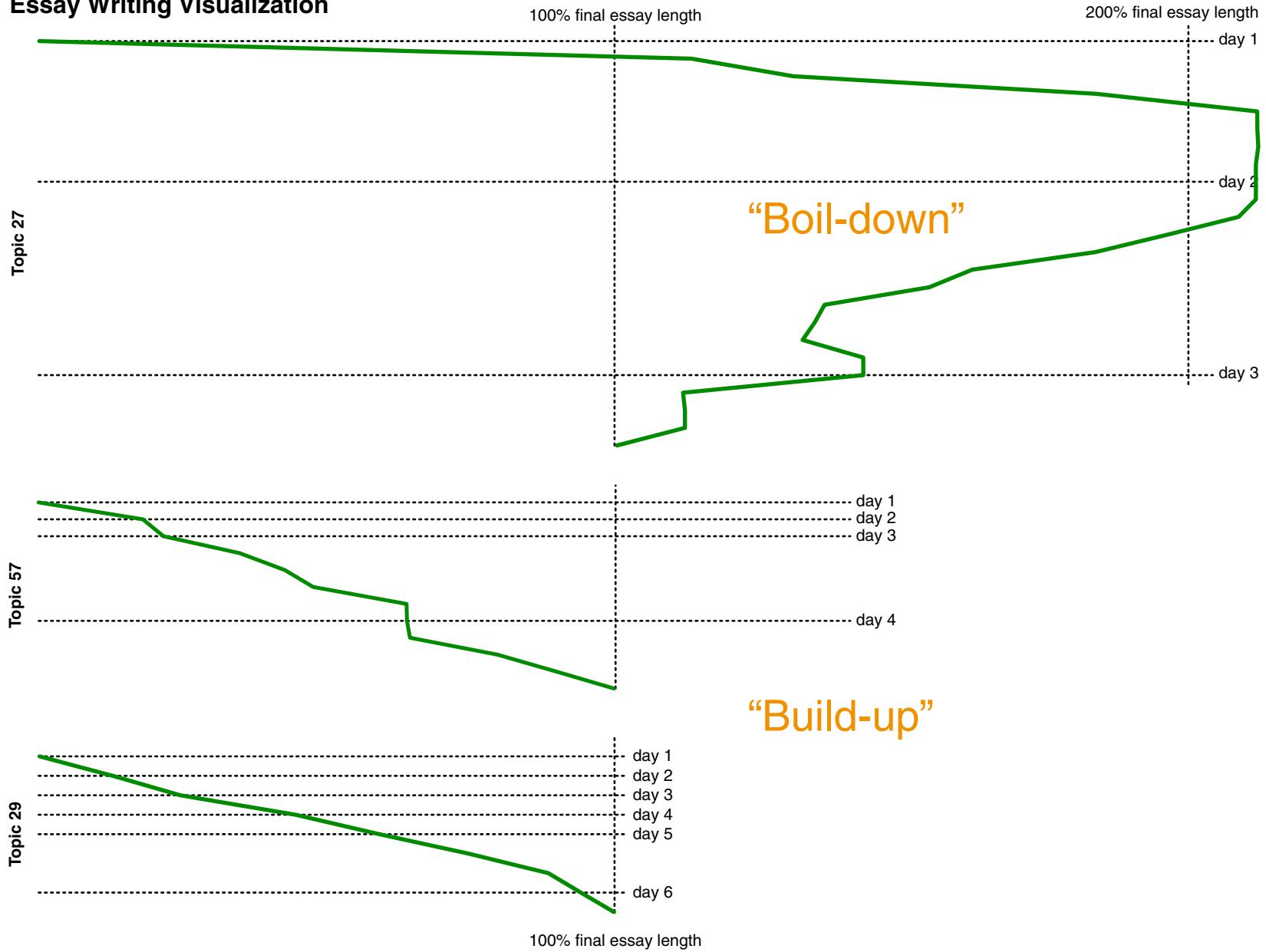
# Essay Writing Visualization



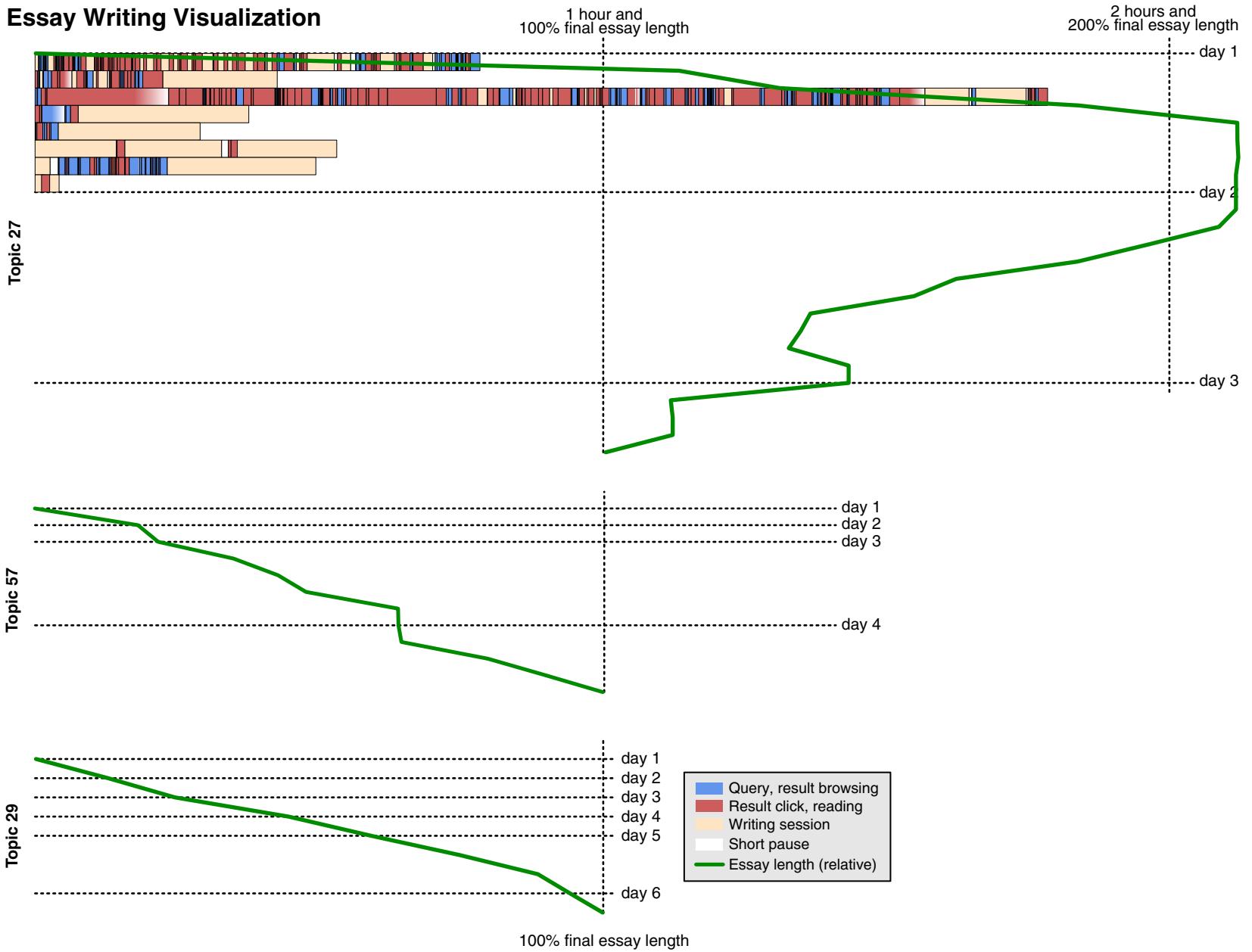
# Essay Writing Visualization



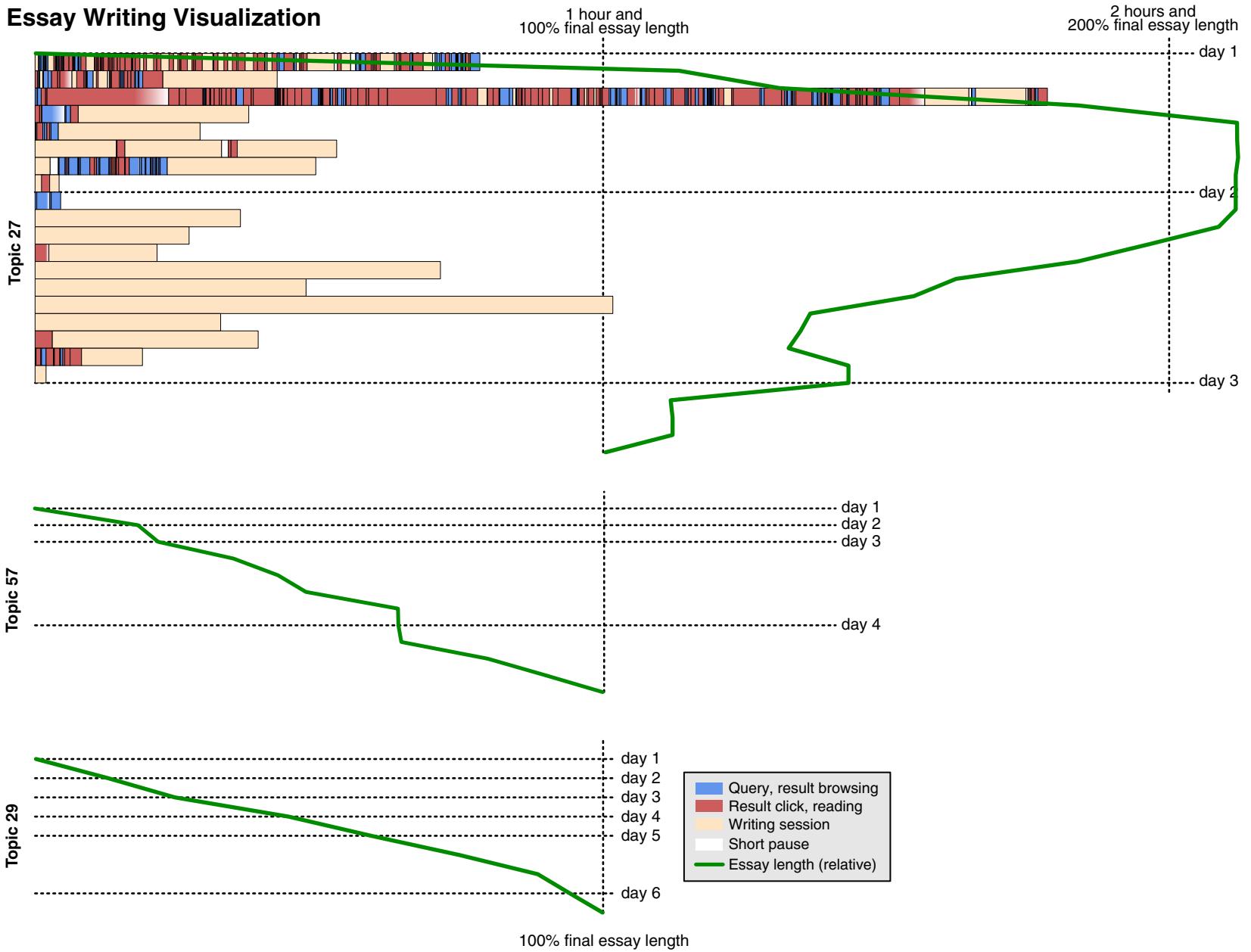
# Essay Writing Visualization



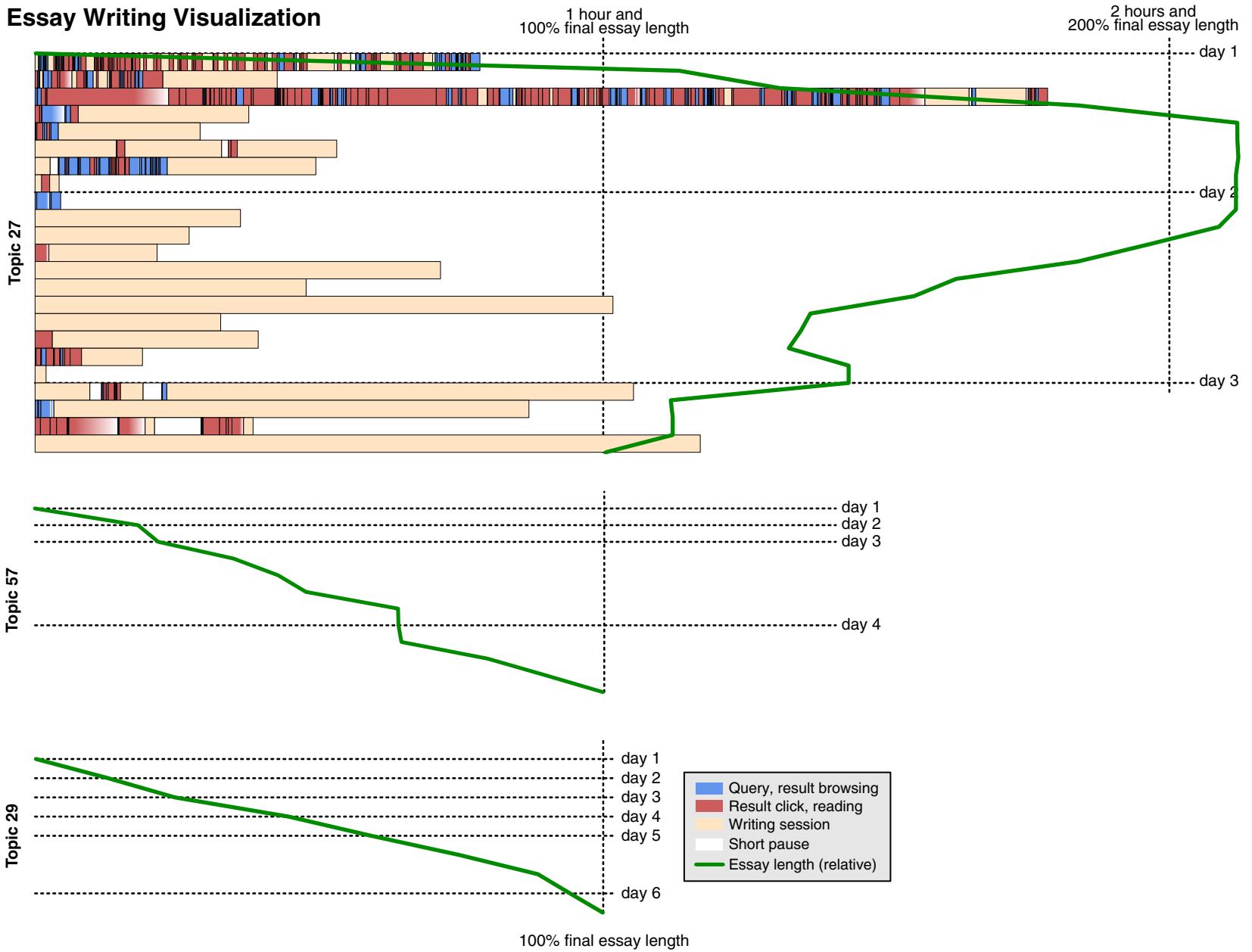
# Essay Writing Visualization



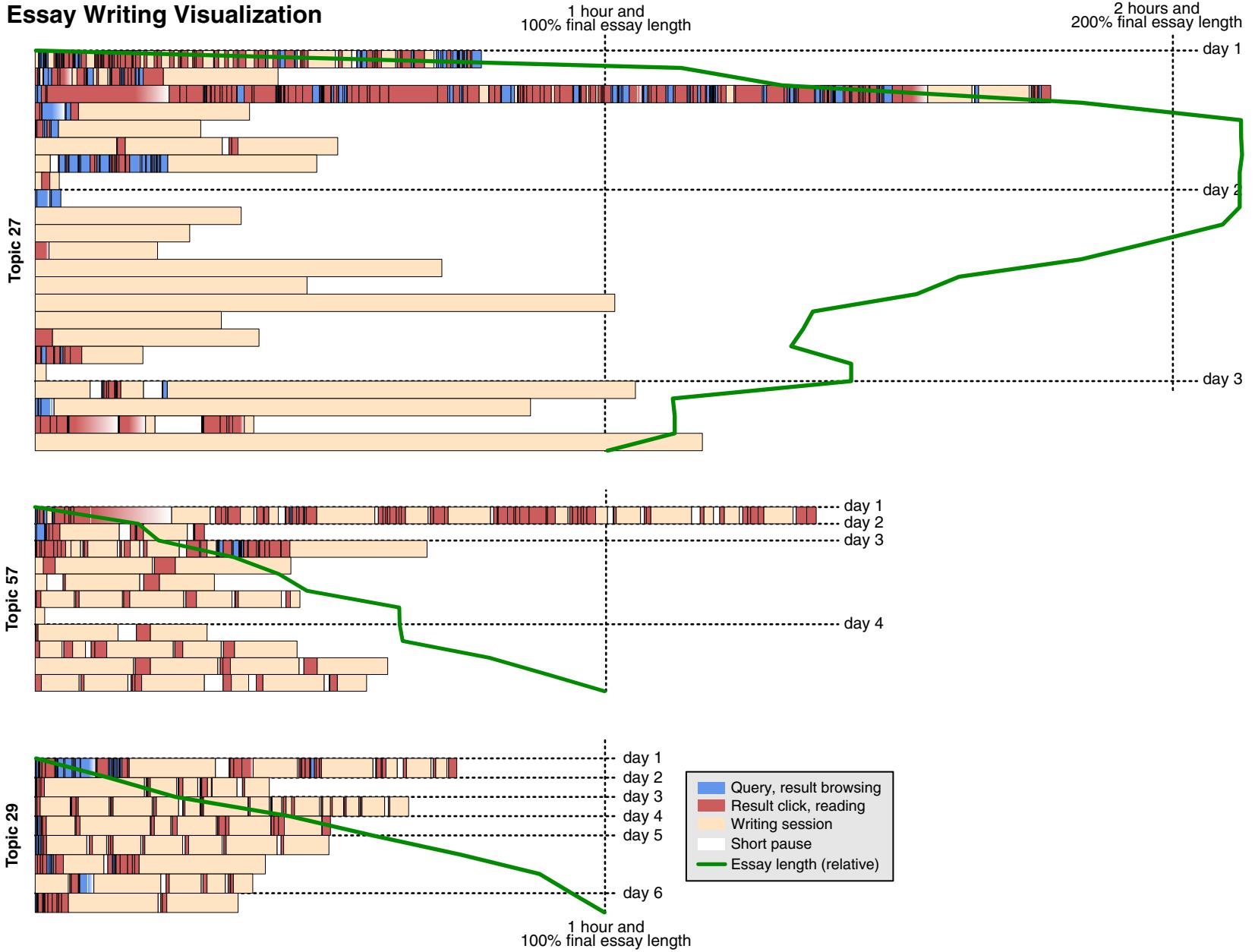
# Essay Writing Visualization



# Essay Writing Visualization



# Essay Writing Visualization

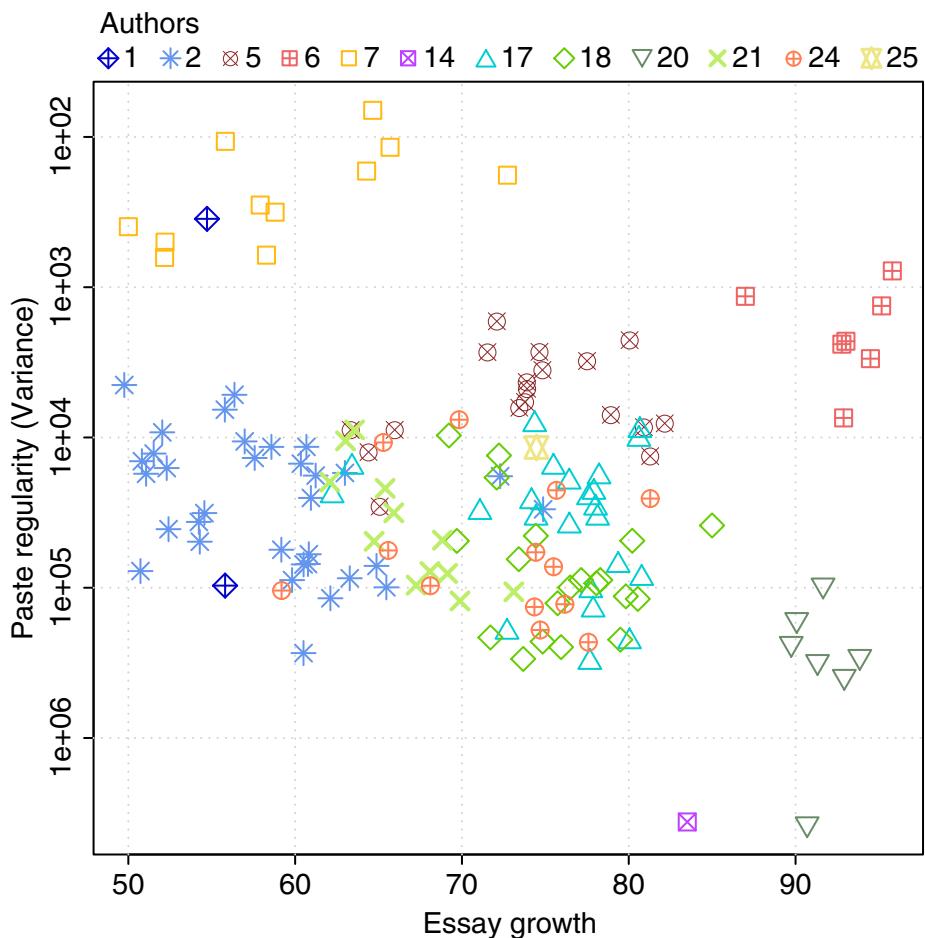


# Writing and Search Behavior

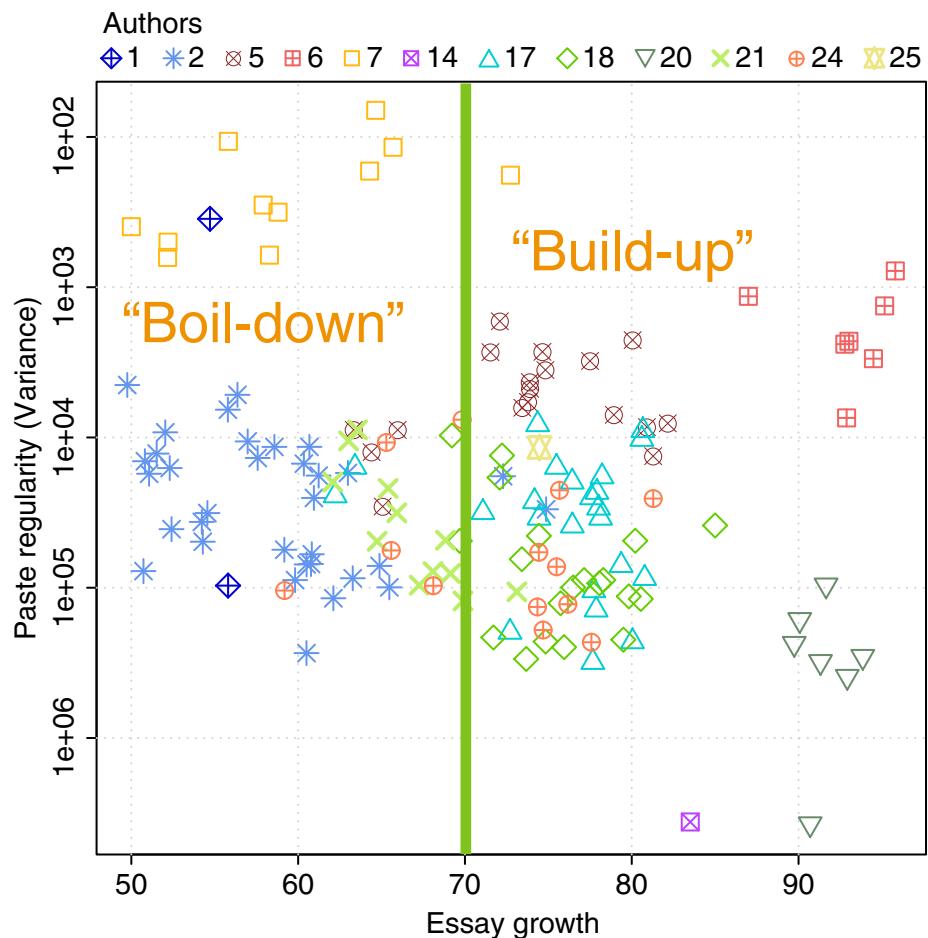
# Writing and Search Behavior

- **Essay Growth:** percentage of lengthening revisions
- **Paste Regularity:** variance of number of revisions between adding sources

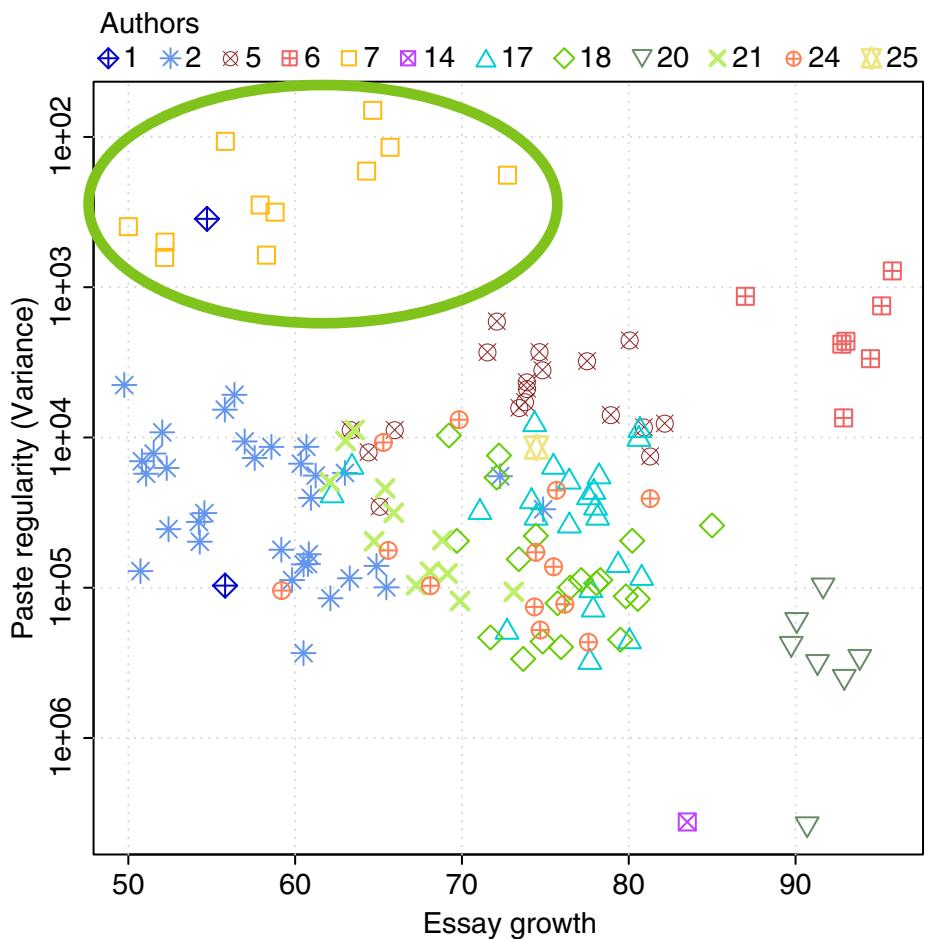
# Writing and Search Behavior



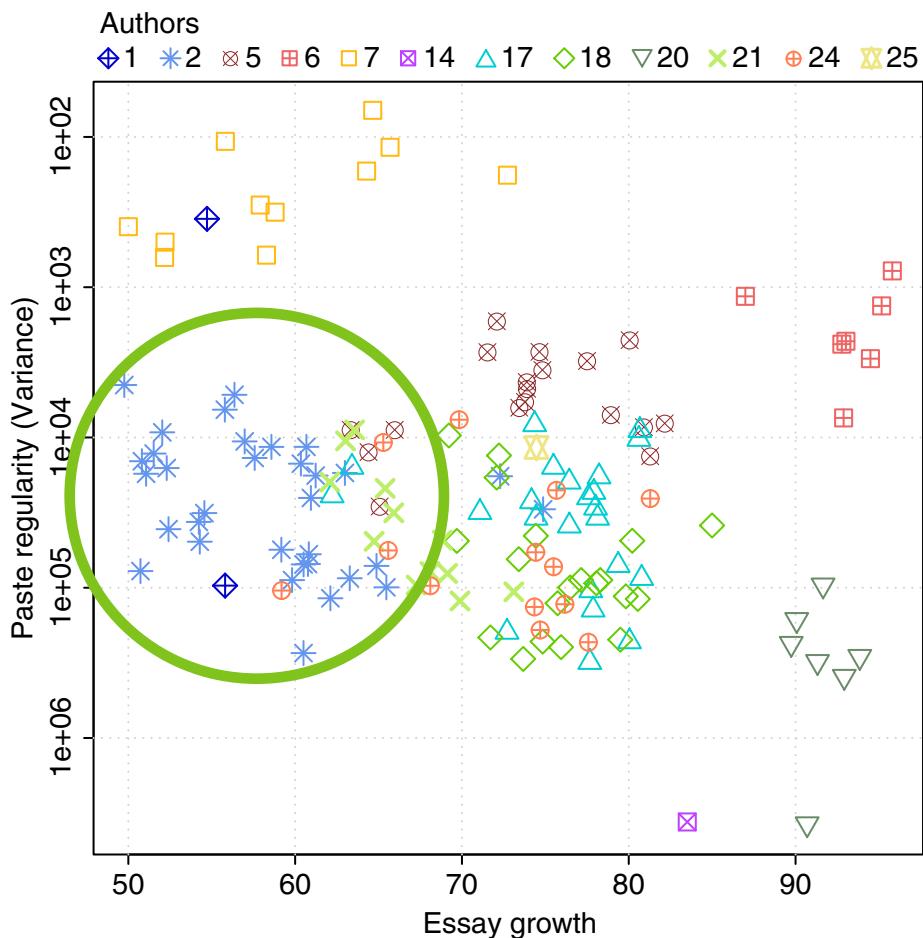
# Writing and Search Behavior



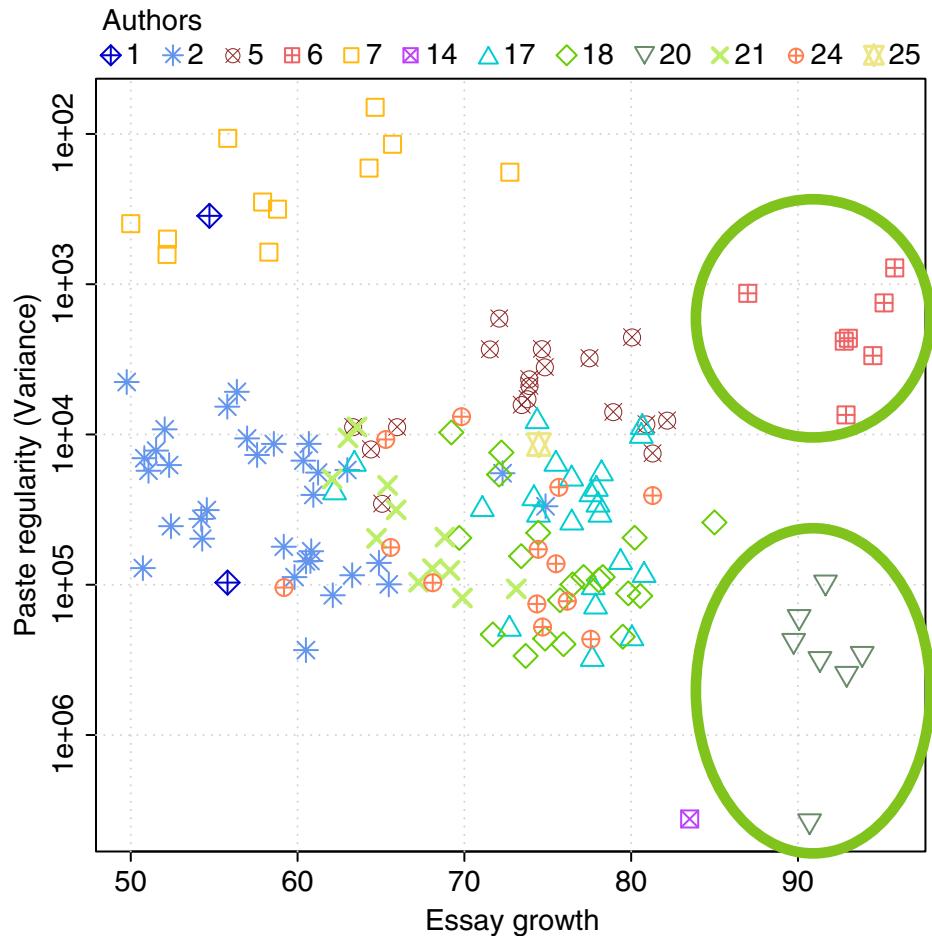
# Writing and Search Behavior



# Writing and Search Behavior



# Writing and Search Behavior



# Query Formulation

# Query Formulation

## Exploratory Search

- Complex, open-ended search task
- Increasing topic understanding
- Iterative refinement

# Origins of New Query Terms

- 3,718 unique query terms

Task description

Search Result Page

Documents

Prior Knowledge

# Origins of New Query Terms

- 3,718 unique query terms

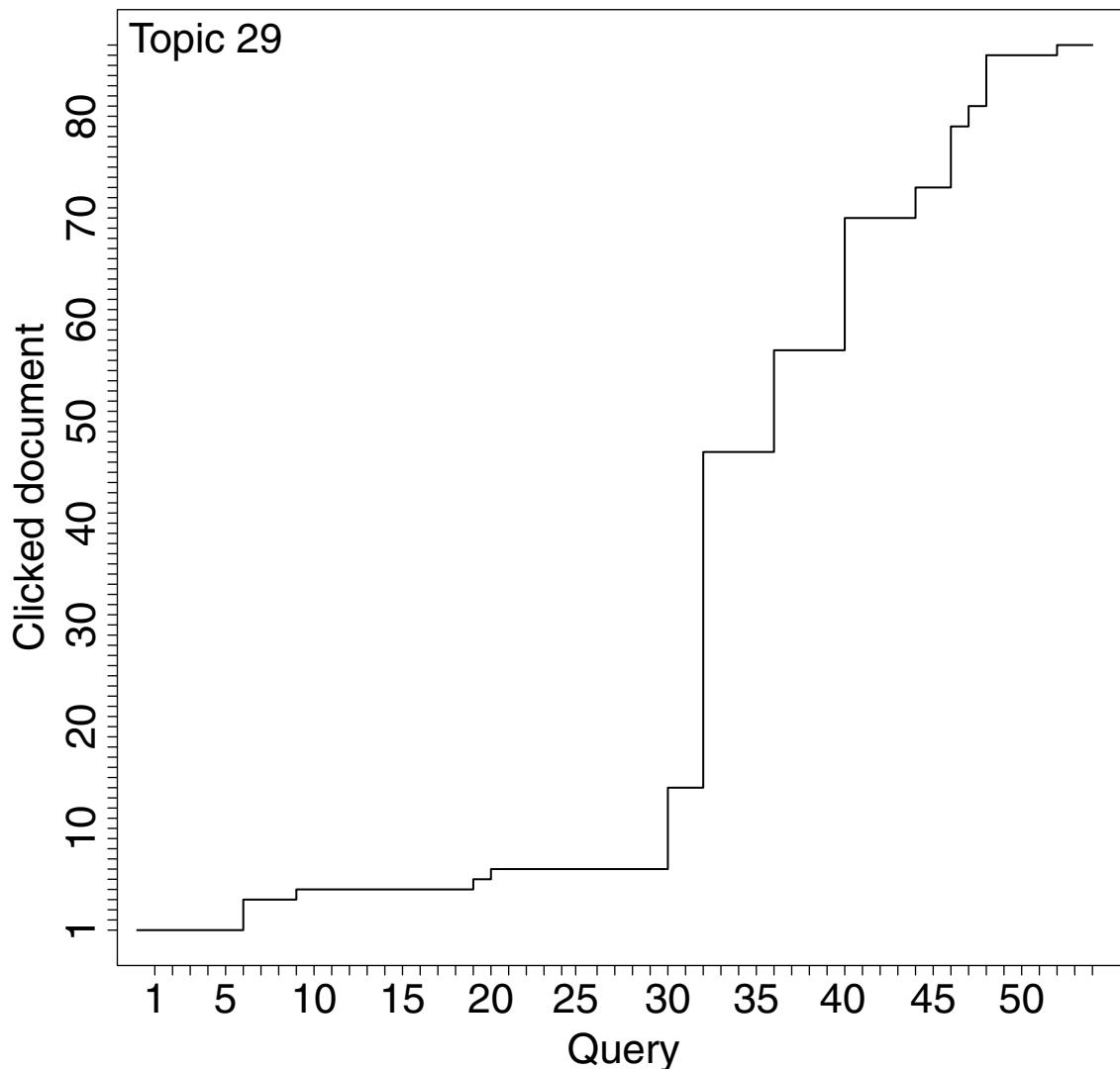
Task description 24%

Search Result Page 37%

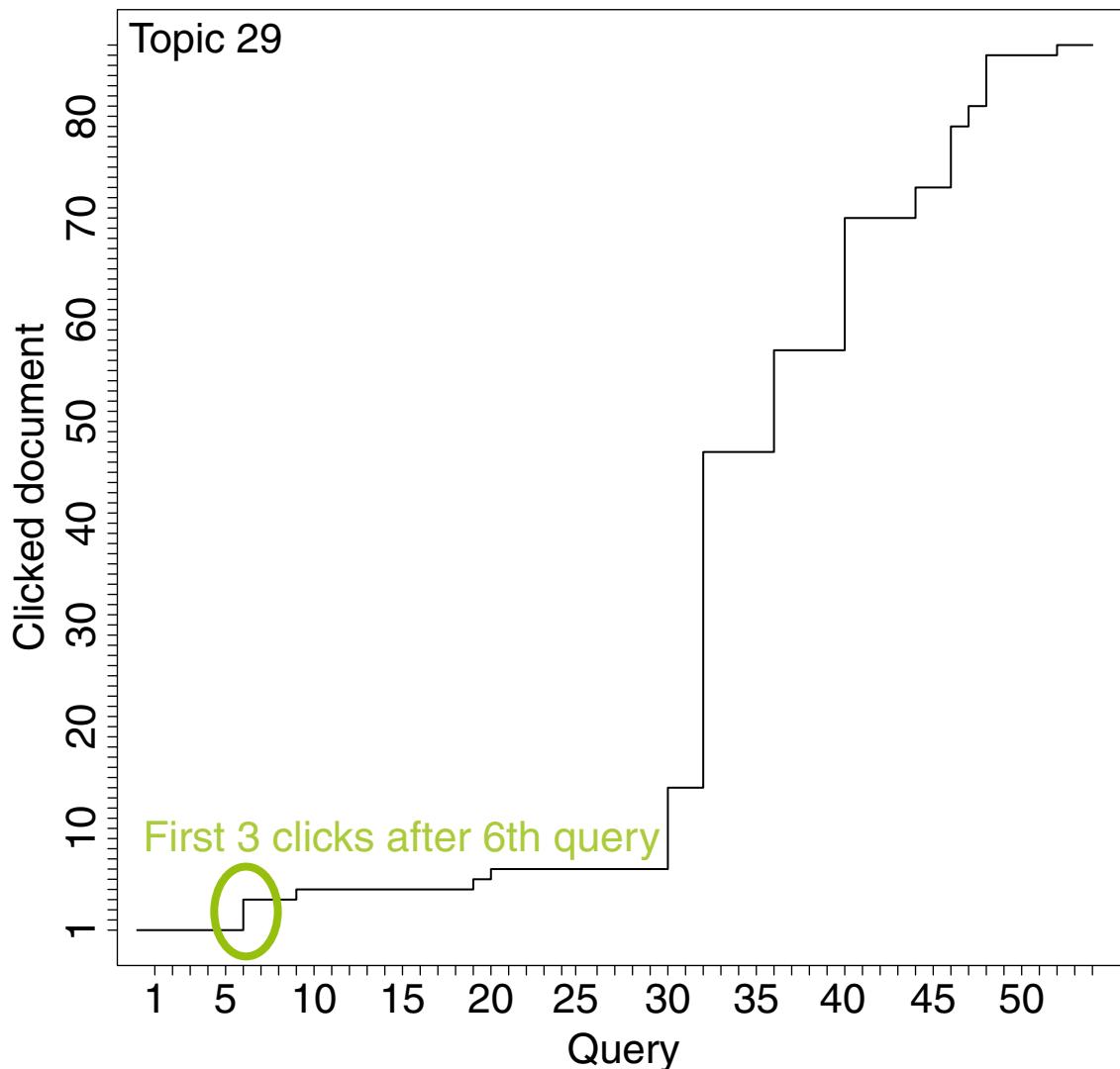
Documents 31%

Prior Knowledge 8%

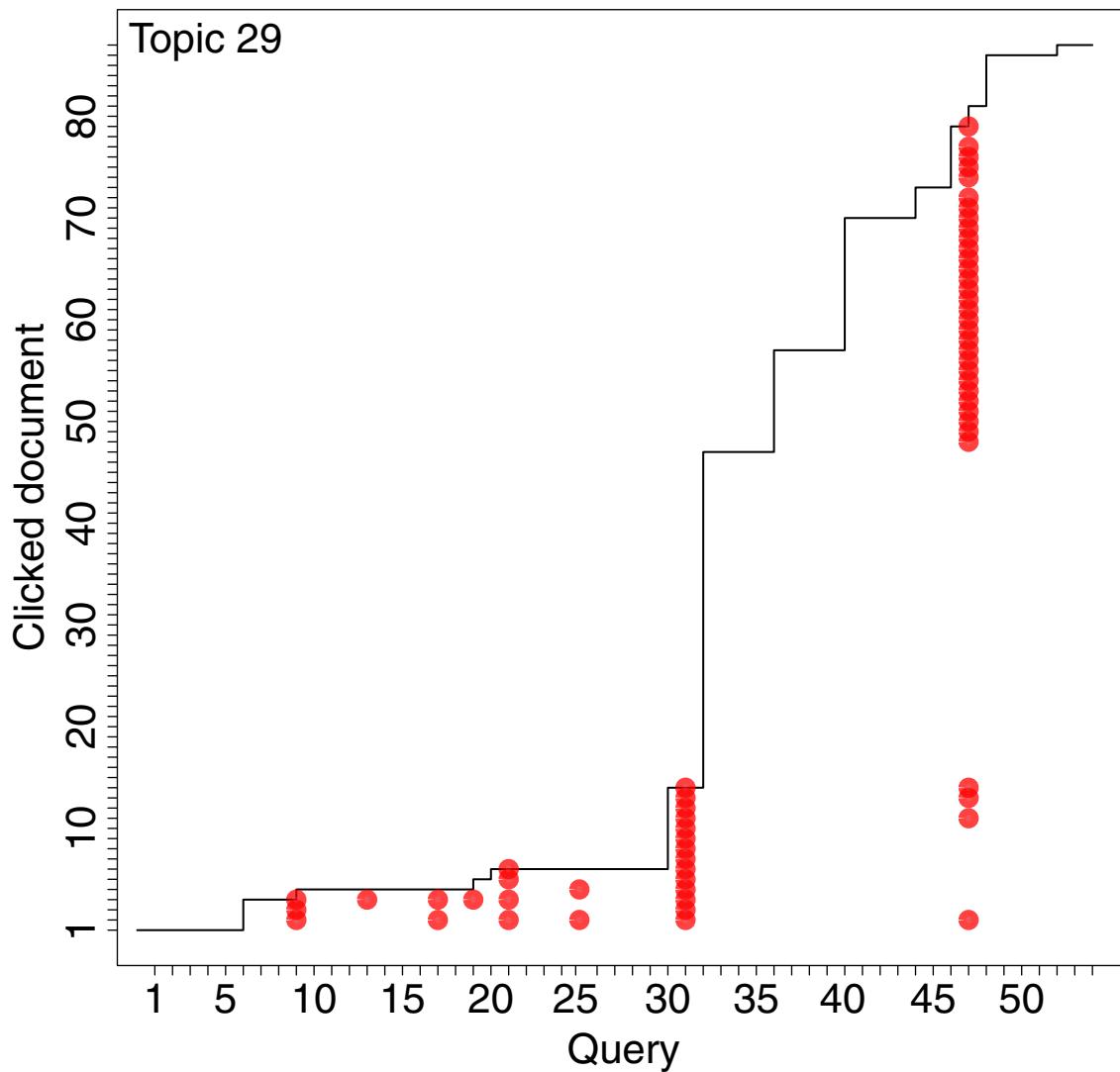
# Origins of New Query Terms



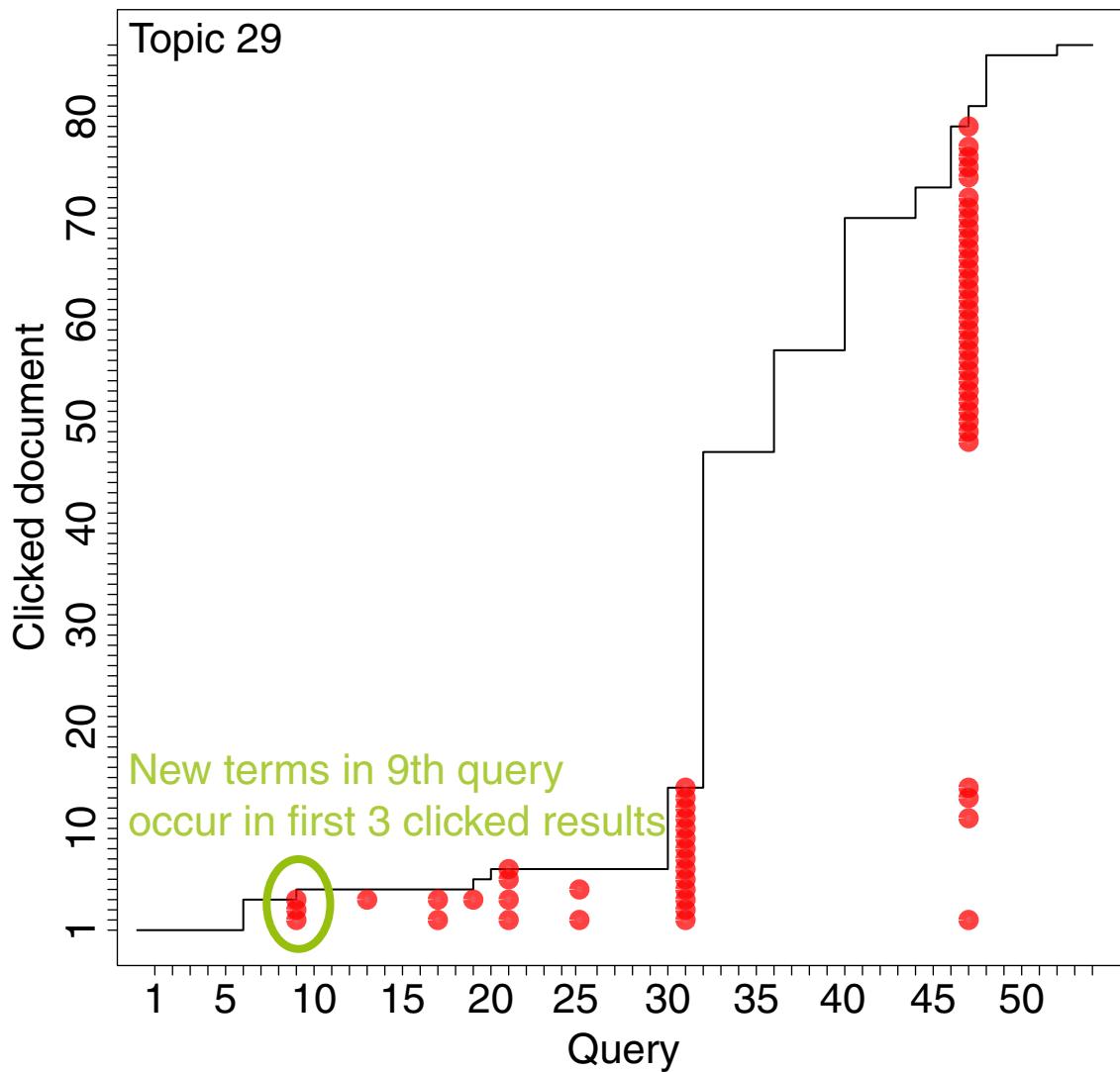
# Origins of New Query Terms



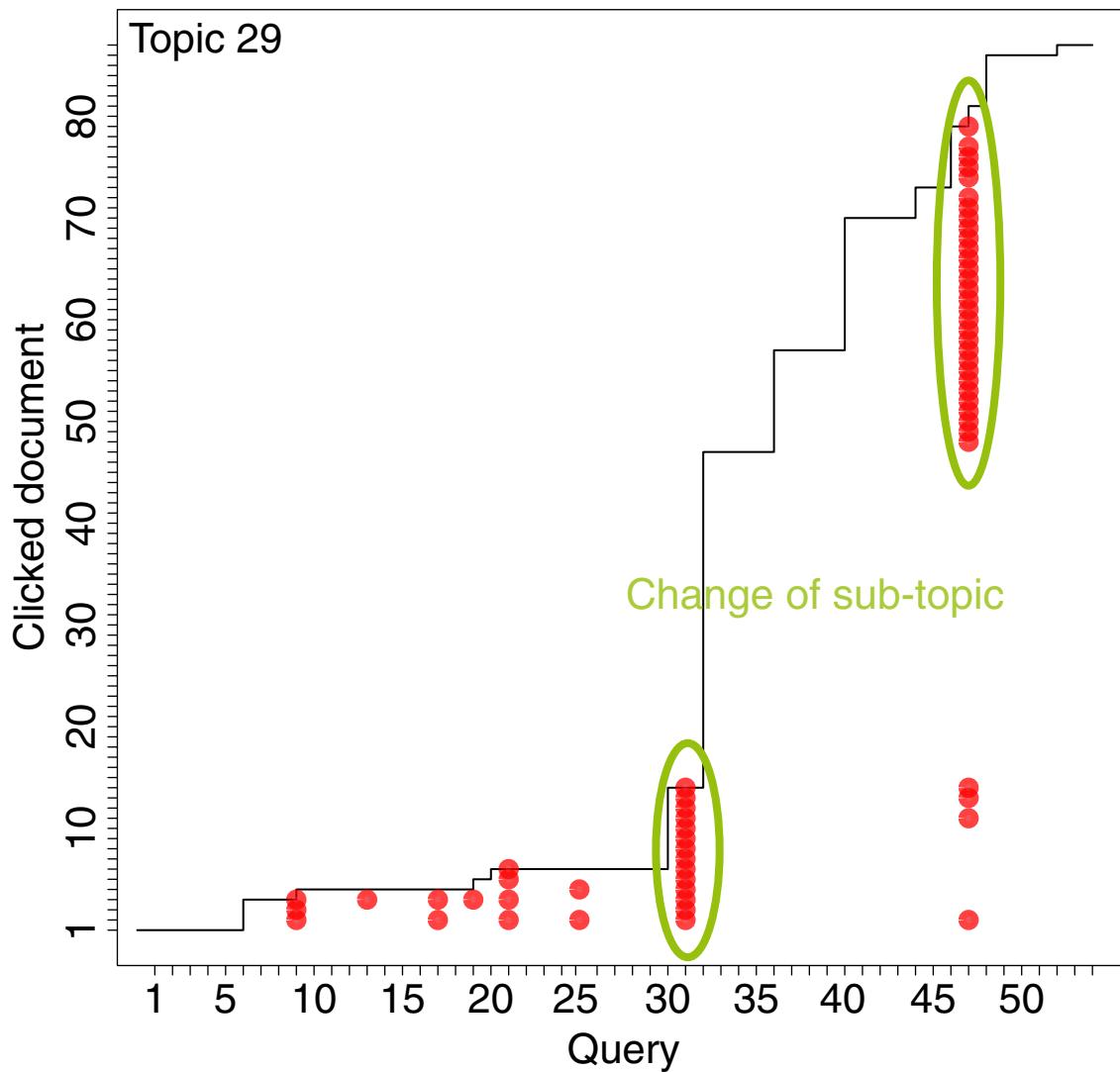
# Origins of New Query Terms



# Origins of New Query Terms



# Origins of New Query Terms



# Summary

# Summary

- A new public dataset to learn how writers search: 150 essays written using web sources, fine-grained revision history and search engine interactions.
- Two distinct searching and writing strategies
- Authors expand subject knowledge while working
- Data is available:  
[www.webis.de](http://www.webis.de) → “Corpora” → “Webis-TRC-12”

# Summary

- A new public dataset to learn how writers search: 150 essays written using web sources, fine-grained revision history and search engine interactions.
- Two distinct searching and writing strategies
- Authors expand subject knowledge while working
- Data is available:  
[www.webis.de](http://www.webis.de) → “Corpora” → “Webis-TRC-12”
- **Thank you!**
- Questions?