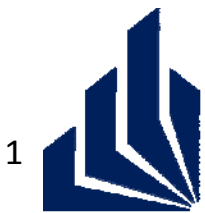


# Vandalism Detection in Wikidata

Stefan Heindorf<sup>1</sup>, Martin Potthast<sup>2</sup>, Benno Stein<sup>2</sup>, Gregor Engels<sup>1</sup>

CIKM 2016

October 25, 2016



**PADERBORN UNIVERSITY**  
*The University for the Information Society*

2

**Bauhaus-Universität Weimar**



About 175,000,000 results (0.57 seconds)

In the news



Obama's response to Trump: 'Come on, man'

CNN - 17 hours ago

... election contests, President **Barack Obama** has boiled his incredulity down to one phrase.

President Barack Obama campaigns for Hillary Clinton in Cleveland

cleveland.com - 18 hours ago

Barack Obama Lays Into Republicans Who Revoked Trump Support: 'You Don't Get Points For That'

Huffington Post - 1 day ago

More news for barack obama

Barack Obama

https://www.barackobama.com/ ▾ Barack Obama ▾

Apr 4, 2016 - OFA works to ensure the voices of ordinary Americans are heard in Washington, while training the next generation of grassroots organizers ...

Barack Obama - Wikipedia

https://en.wikipedia.org/wiki/Barack\_Obama ▾ Wikipedia ▾

**Barack Hussein Obama II** is an American politician who is the 44th and current President of the United States. He is the first African American to hold the office ...

Michelle Obama · Barack Obama, Sr. · Ann Dunham · Early life and career of Barack



Barack Obama



44th U.S. President

Barack Hussein Obama II is an American politician who is the 44th and current President of the United States. He is the first African American to hold the office and the first president born outside the continental United States. [Wikipedia](#)

**Born:** August 4, 1961 (age 55), Kapiolani Medical Center for Women and Children, Honolulu, HI

**Height:** 6' 1"

**Parents:** Ann Dunham, Barack Obama Sr.

**Siblings:** Maya Soetoro-Ng, Auma Obama, David Ndesandjo, More

**Education:** Harvard Law School (1988–1991), More

**Children:** Malia Ann Obama, Sasha Obama

# Motivation



**Barack Obama** 

44th U.S. President

Barack Hussein Obama II is an American politician who is the 44th and current President of the United States. He is the first African American to hold the office and the first president born outside the continental United States. [Wikipedia](#)

**Born:** August 4, 1961 (age 55), Kapiolani Medical Center for Women and Children, Honolulu, HI

**Height:** 6' 1"

**Parents:** [Ann Dunham](#), [Barack Obama Sr.](#)

**Siblings:** [Maya Soetoro-Ng](#), [Auma Obama](#), [David Ndesandjo](#), [More](#)

**Education:** [Harvard Law School \(1988–1991\)](#), [More](#)

**Children:** [Malia Ann Obama](#), [Sasha Obama](#)

# Motivation



A screenshot of the top portion of the Barack Obama Wikipedia article. It features a grid of images of Barack Obama, his name "Barack Obama" in a large font, and his title "44th U.S. President". Below this is a short biographical paragraph, followed by key facts such as "Born: August 4, 1961", "Height: 6' 1\"", "Parents: Ann Dunham, Barack Obama Sr.", "Siblings: Maya Soetoro-Ng, Auma Obama, David Ndesandjo", "Education: Harvard Law School (1988–1991)", and "Children: Malia Ann Obama, Sasha Obama".

# Motivation



# Motivation



### Barack Obama



Obama in front of the *Resolute* desk in the [Oval Office](#) of the [White House](#) on December 6, 2012

**44th President of the United States**

**Incumbent**

**Assumed office**  
January 20, 2009

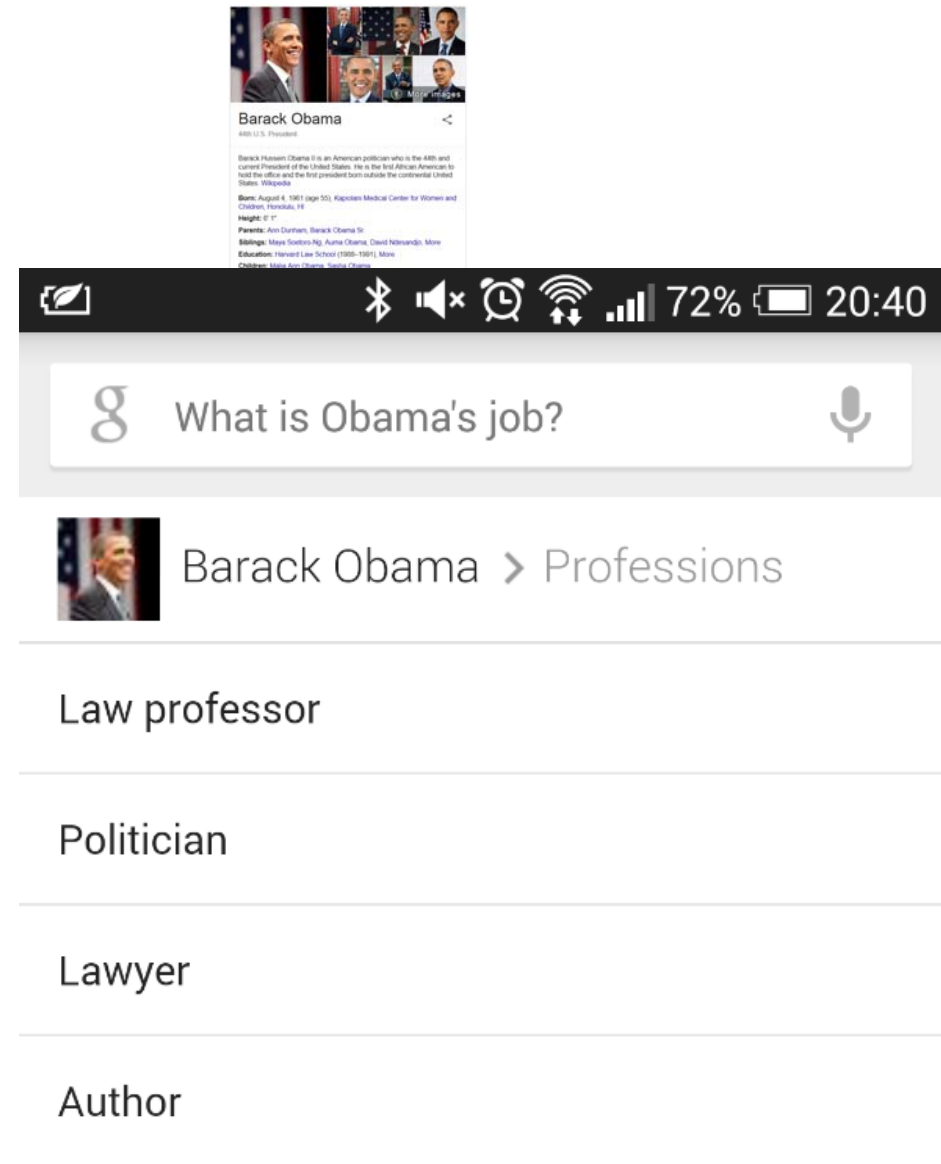
**Vice President** [Joe Biden](#)

**Preceded by** [George W. Bush](#)

# Motivation



# Motivation





# Motivation



# Barack Obama (Q76)

Worst president ever

 edit

## Statements

sex or gender

 shemale

 edit

▸ 2 references

+ add

spouse

 Peter Piper

 edit

▸ 0 references

+ add

# Barack Obama (Q76)

Worst president ever

 edit

## Statements

sex or gender

 shemale

 edit

▸ 2 references

+ add

spouse

 Peter Piper

 edit


▸ 0 references

+ add



} Item head

**Barack Obama** (Q76)



---

Worst president ever  edit

**Statements**

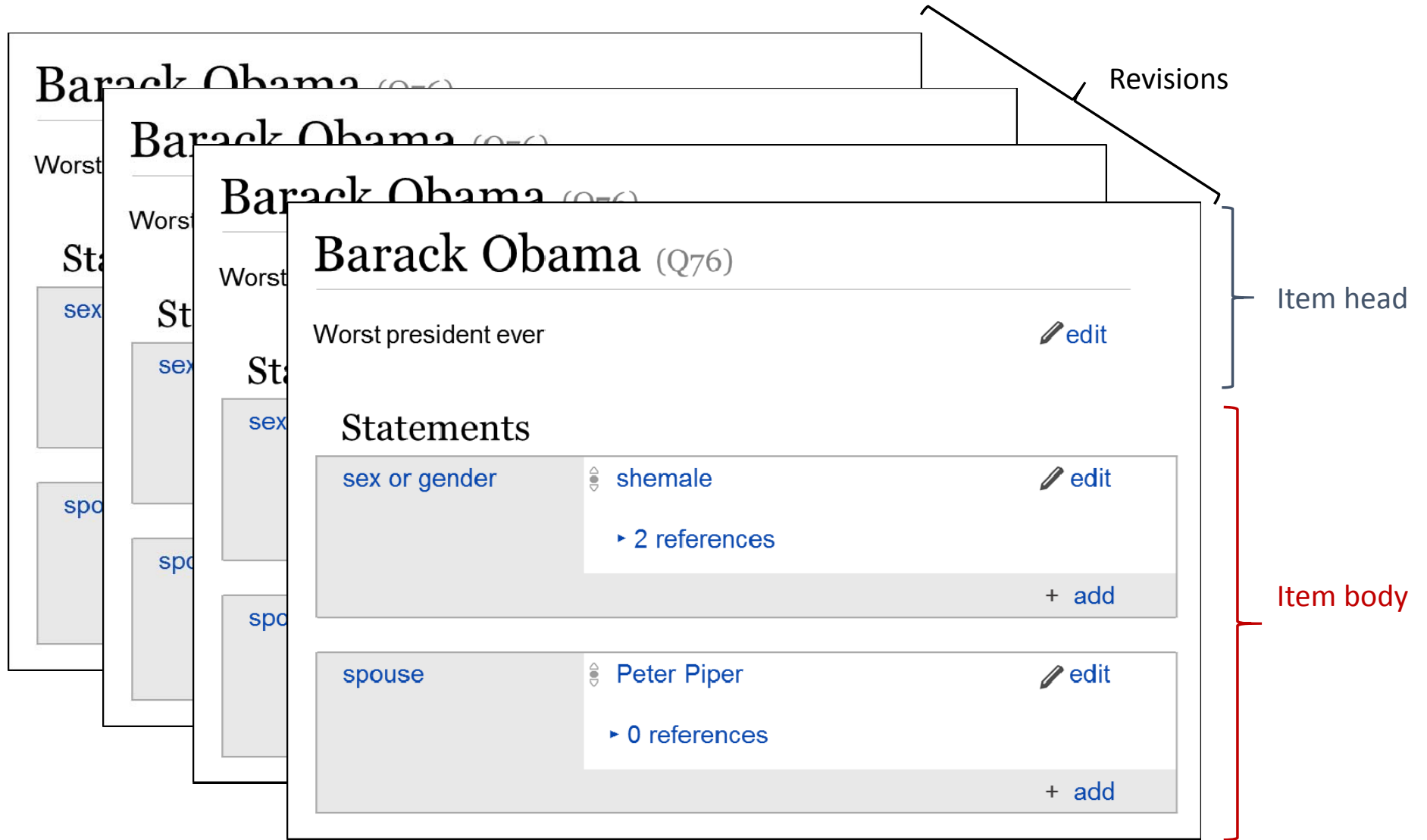
sex or gender	 shemale <span style="float: right;"> edit</span>
	▶ 2 references
	+ add

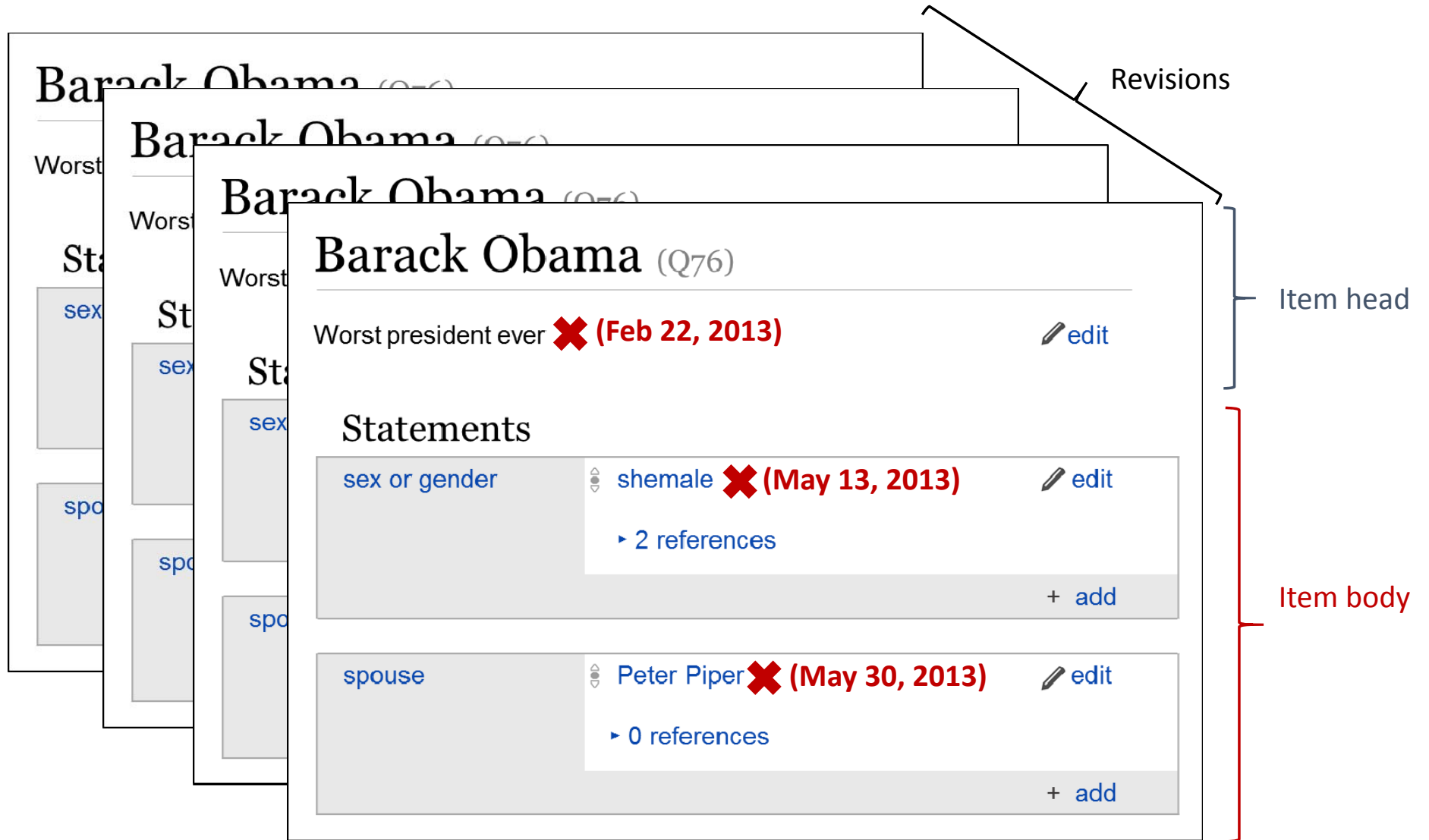
  

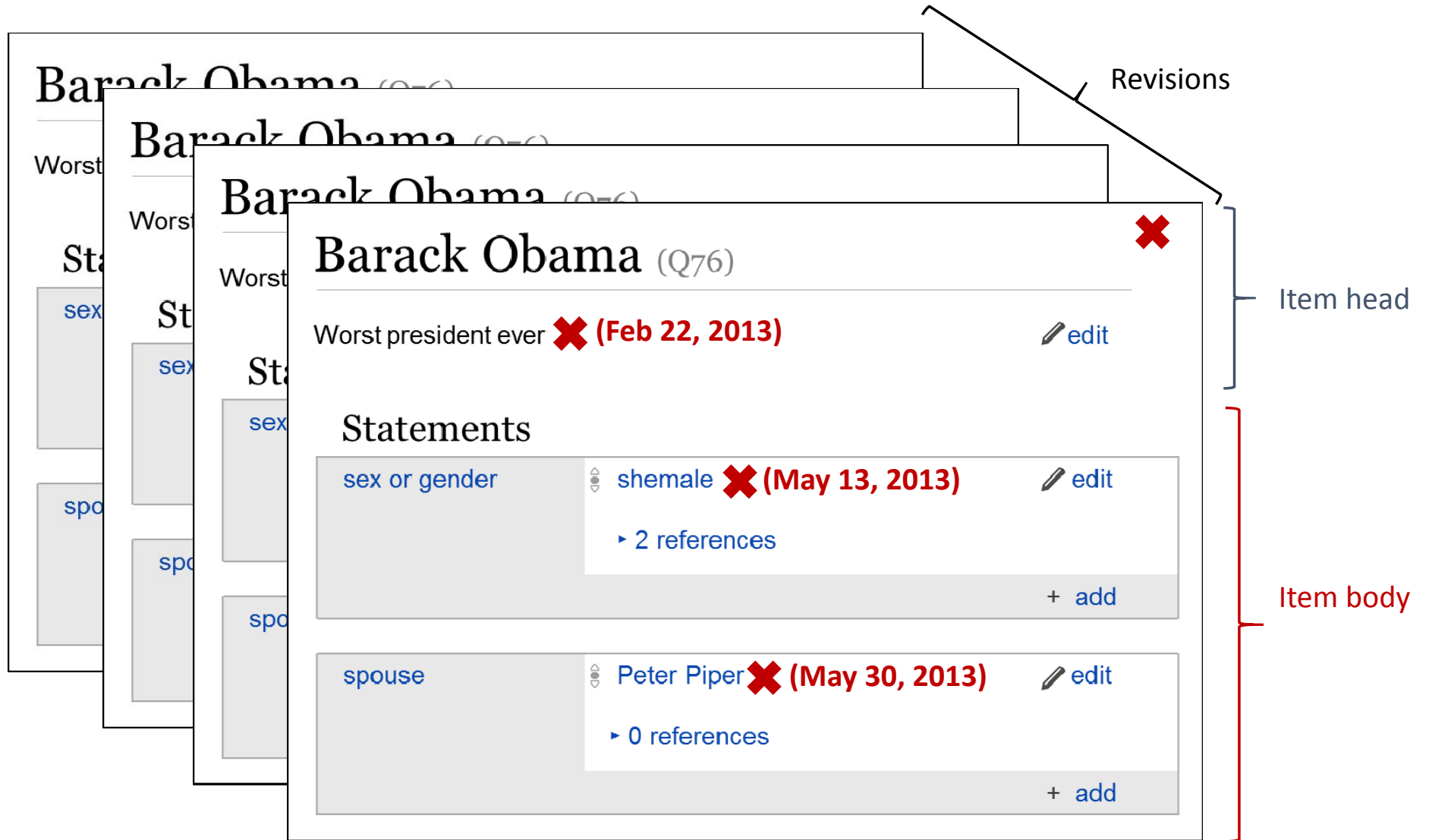
spouse	 Peter Piper <span style="float: right;"> edit</span>
	▶ 0 references
	+ add

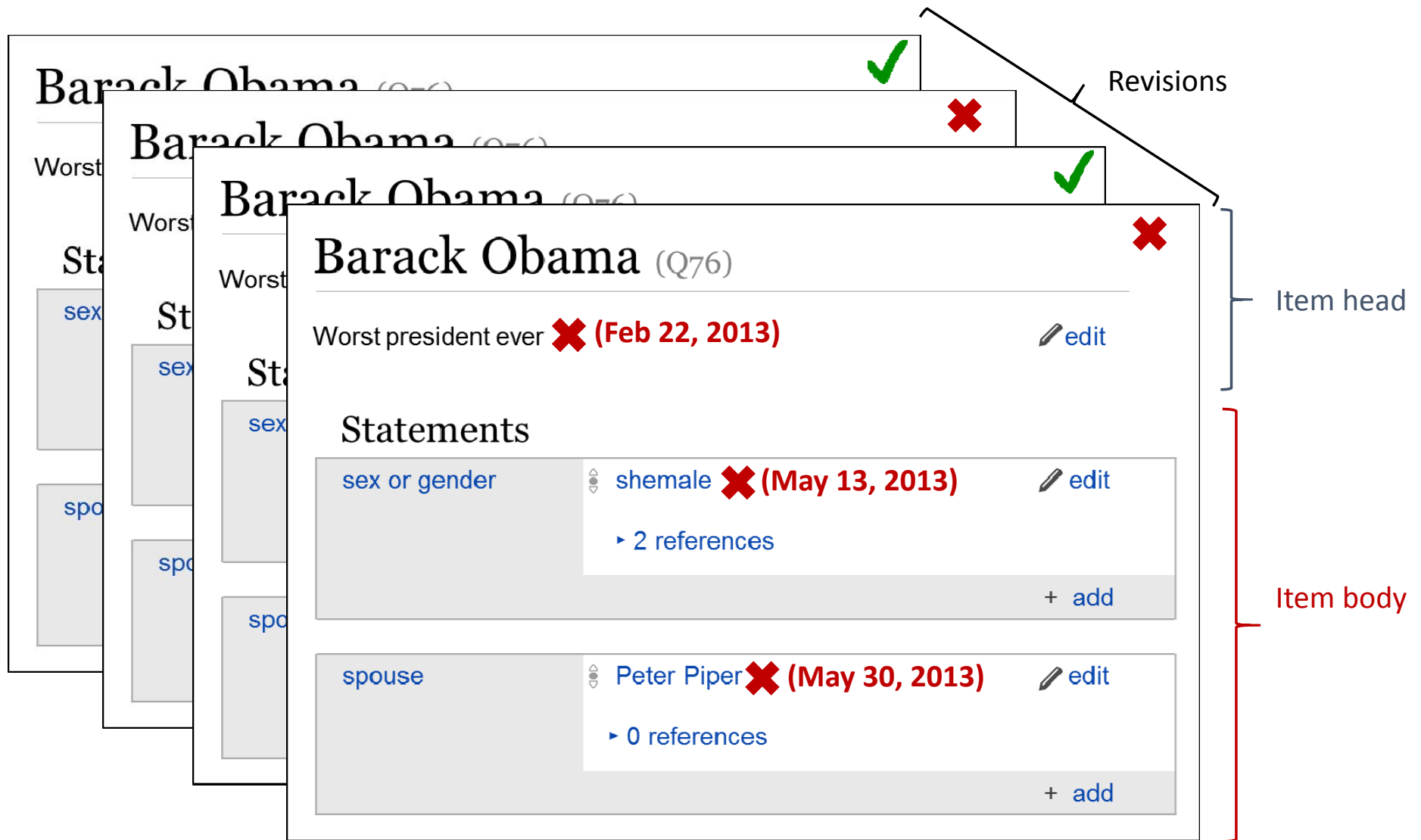
} Item head

} Item body











Barack Obama (Q76)

Worst president ever ✖ (Feb 22, 2013) [edit](#)

Statements

sex or gender [shemale ✖ \(May 13, 2013\)](#) [edit](#)  
+ 2 references  
[+ add](#)

spouse [Peter Piper ✖ \(May 30, 2013\)](#) [edit](#)  
▸ 0 references  
[+ add](#)

Revisions

Item head

Item body

**Vandalism**

# Why is it a problem?

Patrolling



Reverting



Warning



Blocking



Protecting



- Over 2 Mio manual edits per month
- A lot of tedious work
- Vandalism is not detected in time

# Research Question

How to detect damaging changes to  
crowdsourced knowledge bases?

# Our Approach

# Our Approach

1. Label Dataset

→ Vandalism Corpus [SIGIR'15]

# Our Approach

1. Label Dataset → Vandalism Corpus [SIGIR'15]
2. Study Vandalism Characteristics → 47 Features

# Our Approach

1. Label Dataset → Vandalism Corpus [SIGIR'15]
2. Study Vandalism Characteristics → 47 Features
3. Experiment with ML → Multiple-Instance Learning

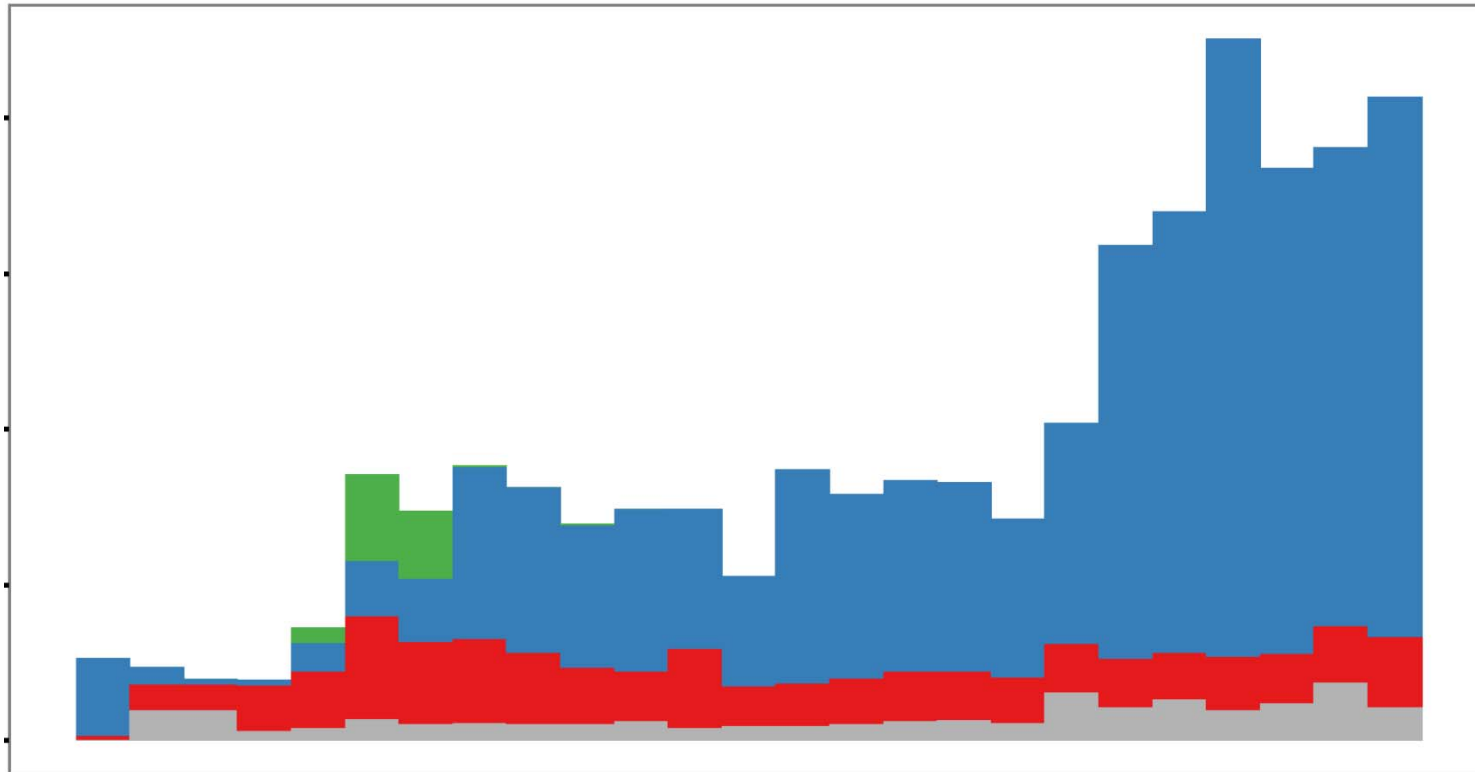
# Our Approach

1. Label Dataset → Vandalism Corpus [SIGIR'15]
2. Study Vandalism Characteristics → 47 Features
3. Experiment with ML → Multiple-Instance Learning
4. Compare with state of the art → 2 Baselines



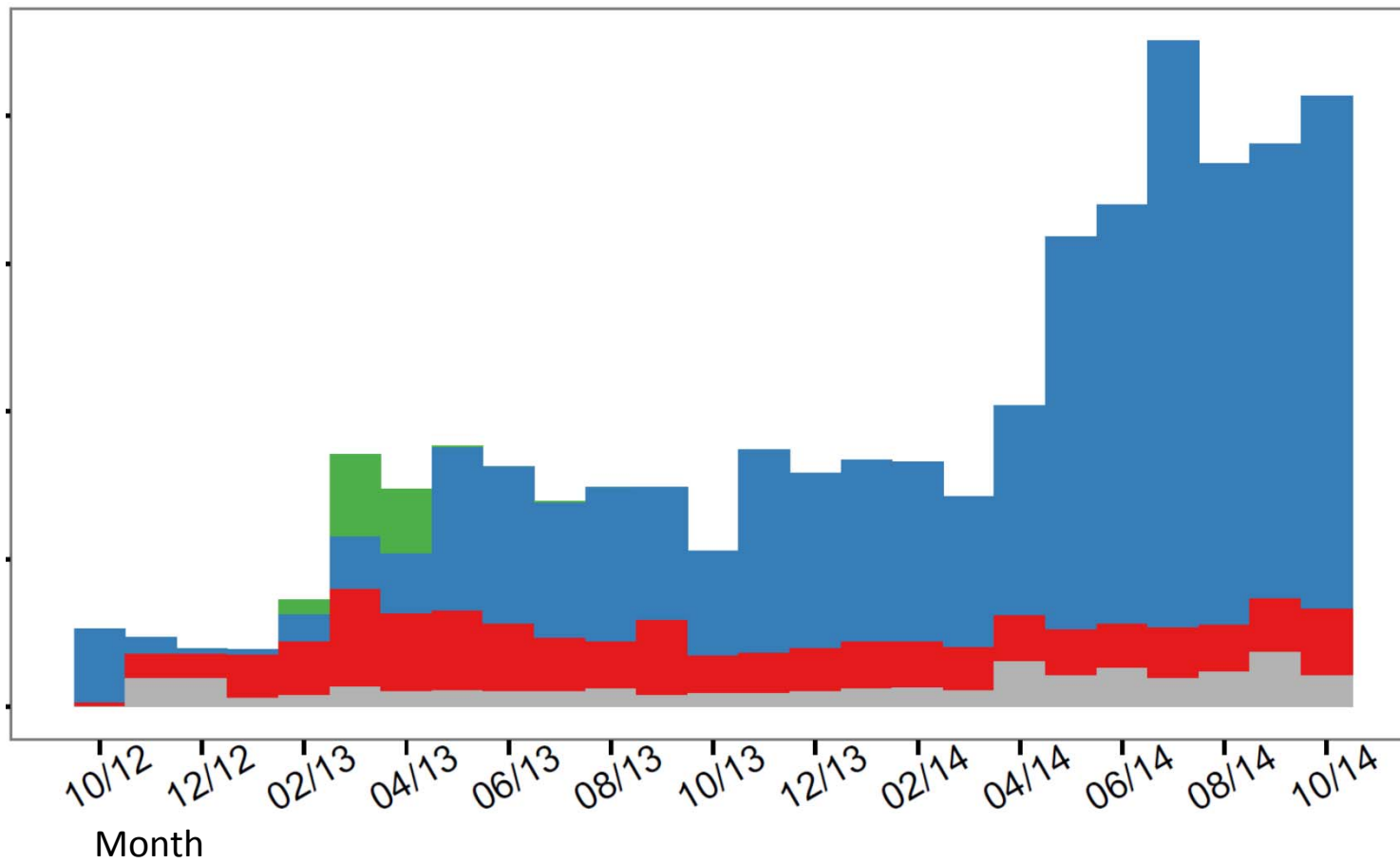
# Corpus [SIGIR '15]

Revisions over time



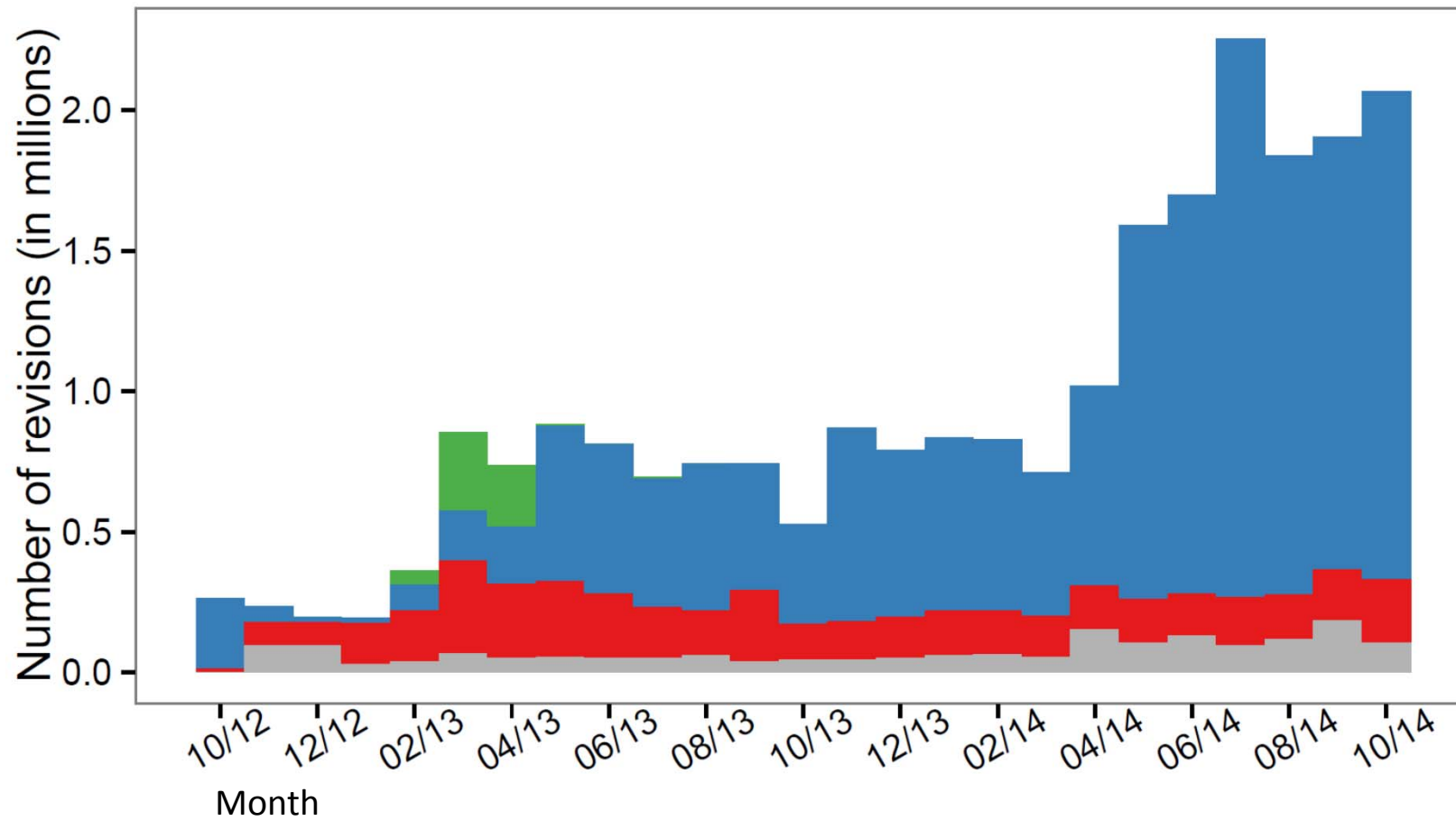
# Corpus [SIGIR '15]

Revisions over time



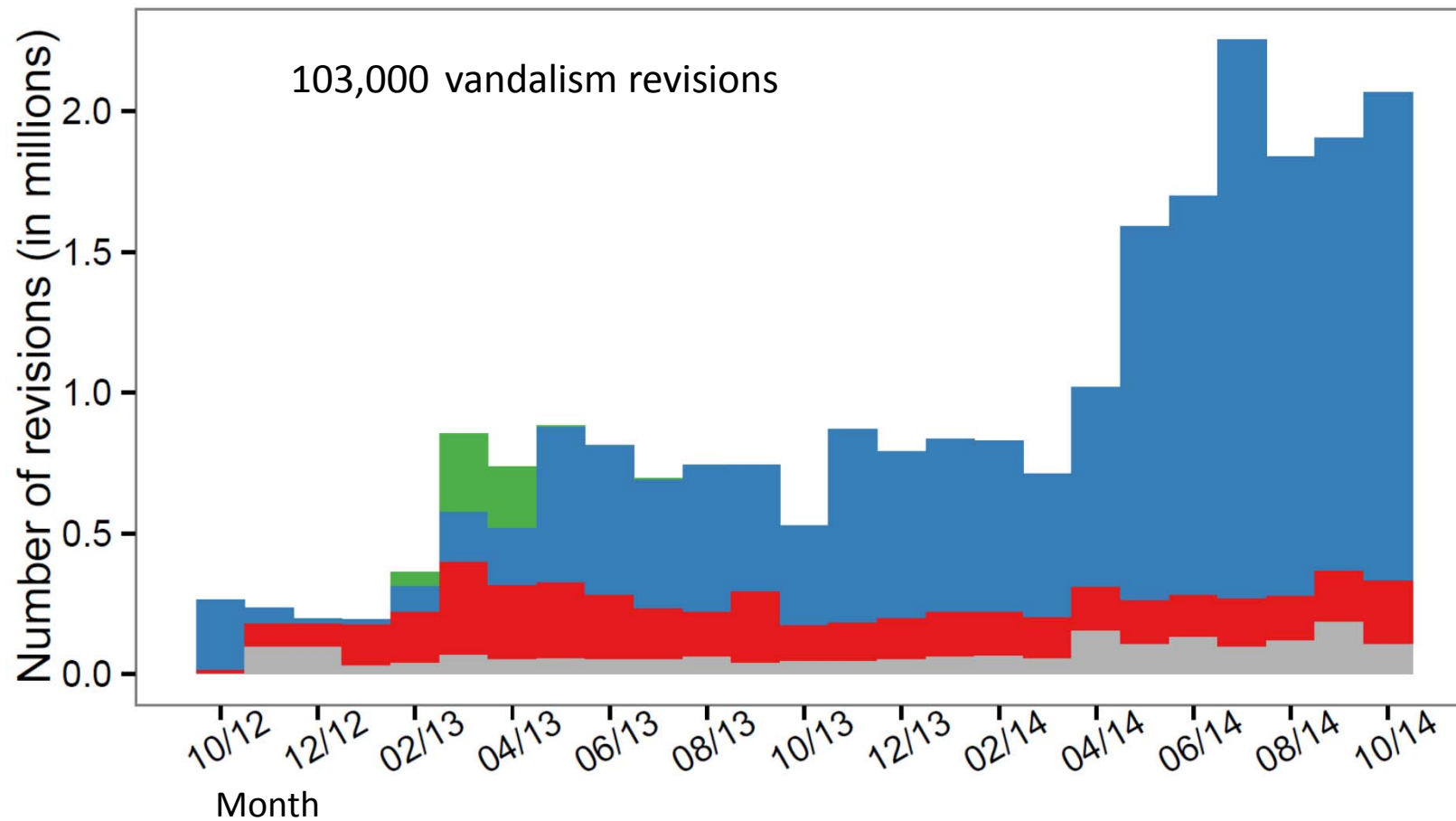
# Corpus [SIGIR '15]

Revisions over time



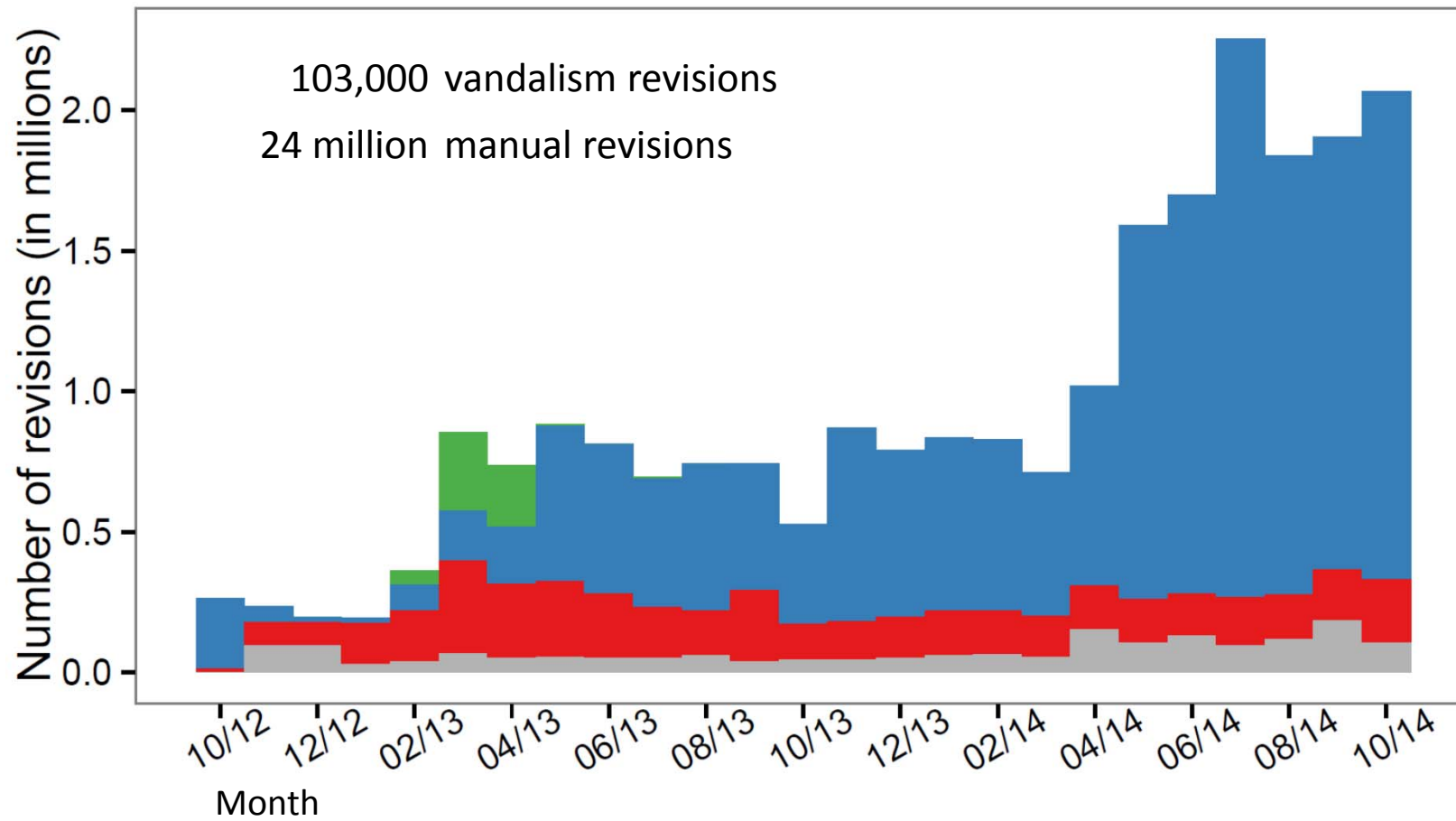
# Corpus [SIGIR '15]

Revisions over time



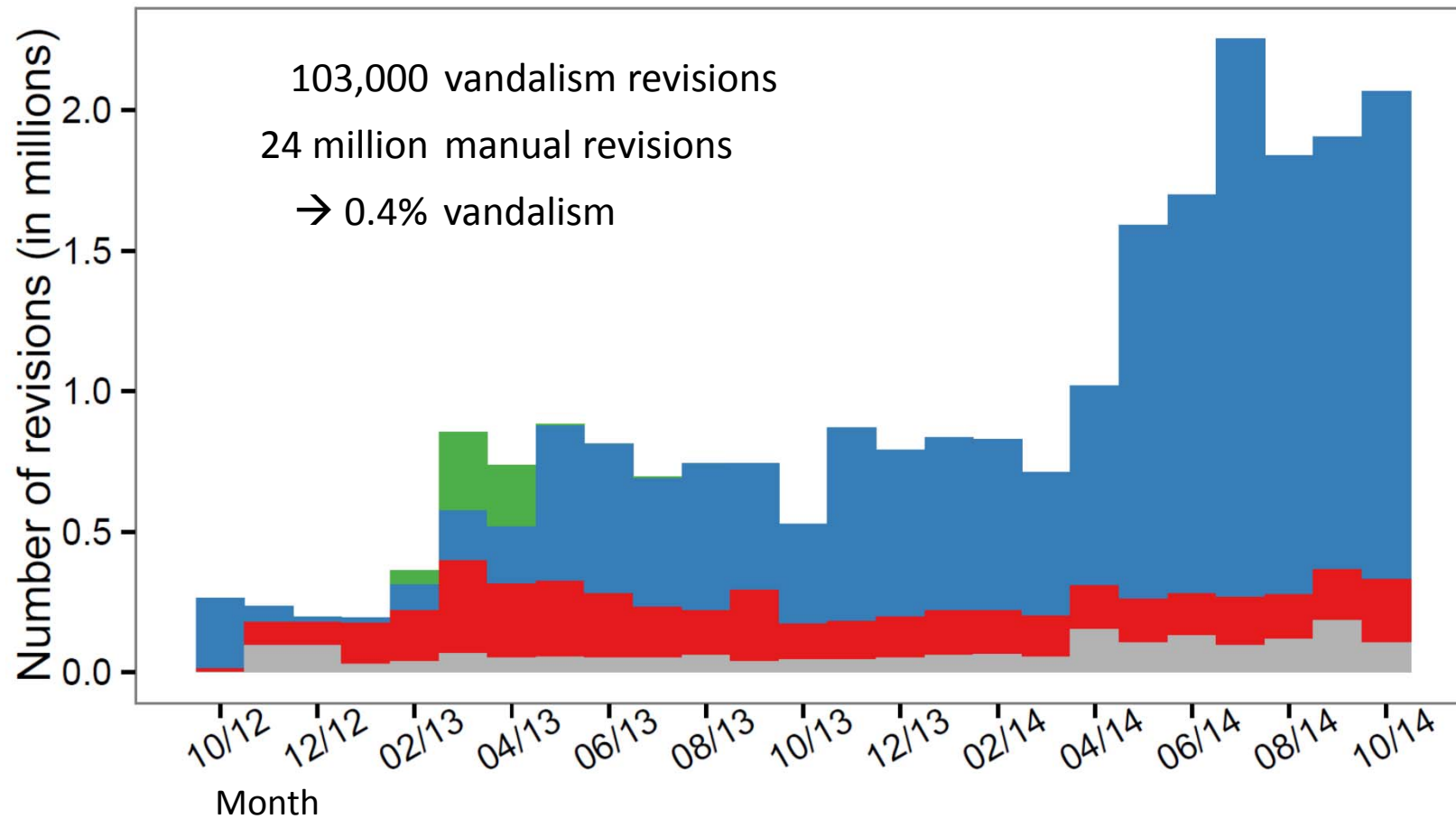
# Corpus [SIGIR '15]

Revisions over time



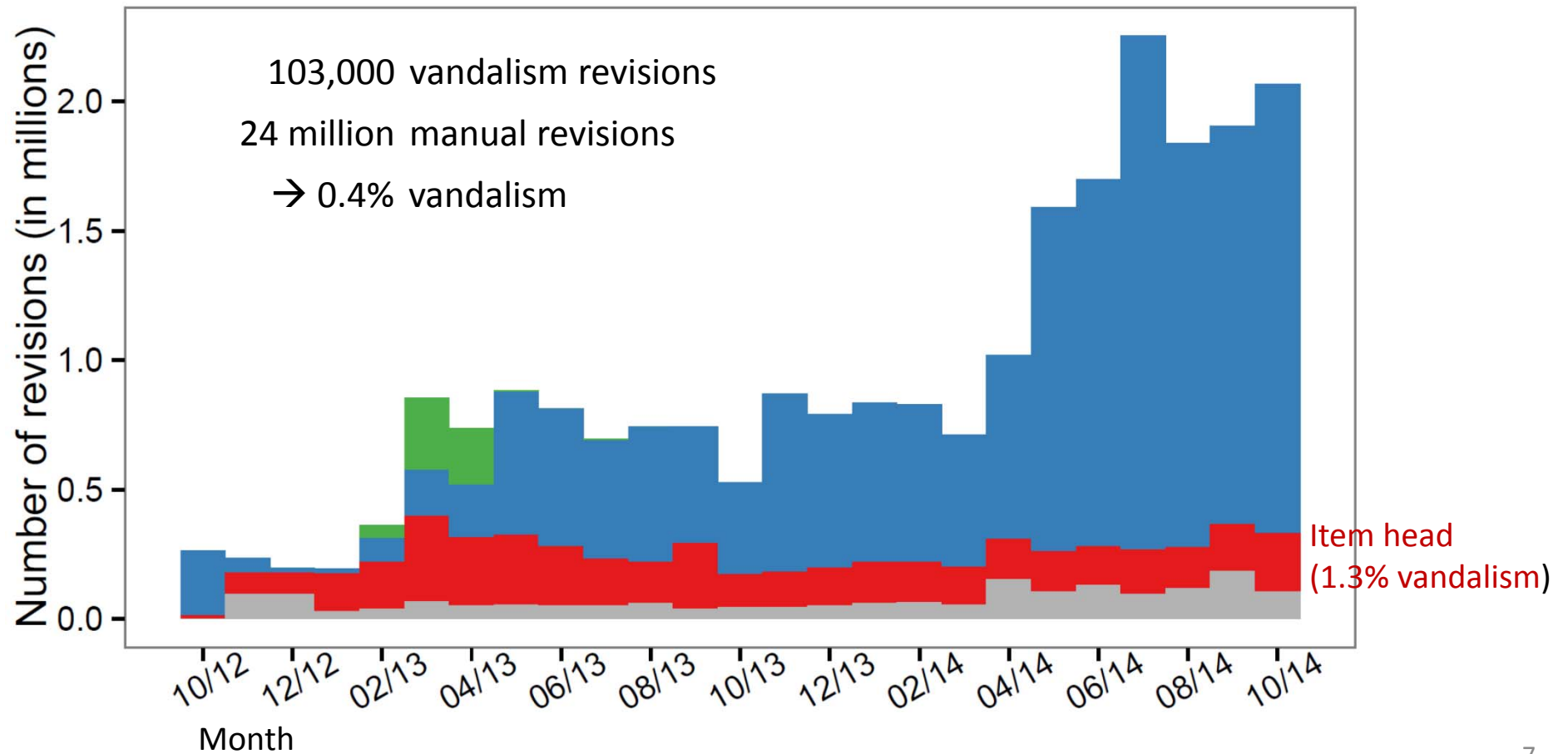
# Corpus [SIGIR '15]

Revisions over time



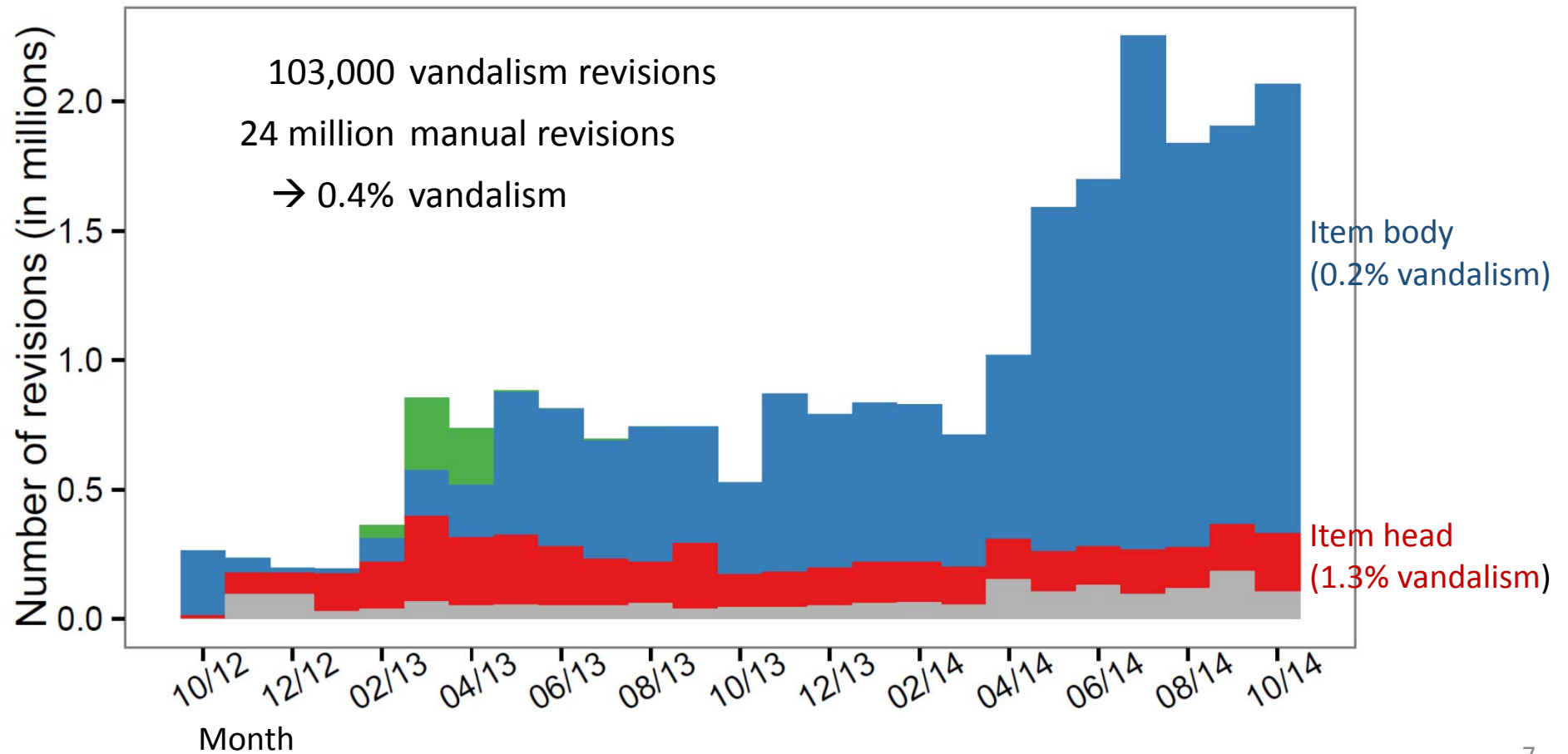
# Corpus [SIGIR '15]

## Revisions over time



# Corpus [SIGIR '15]

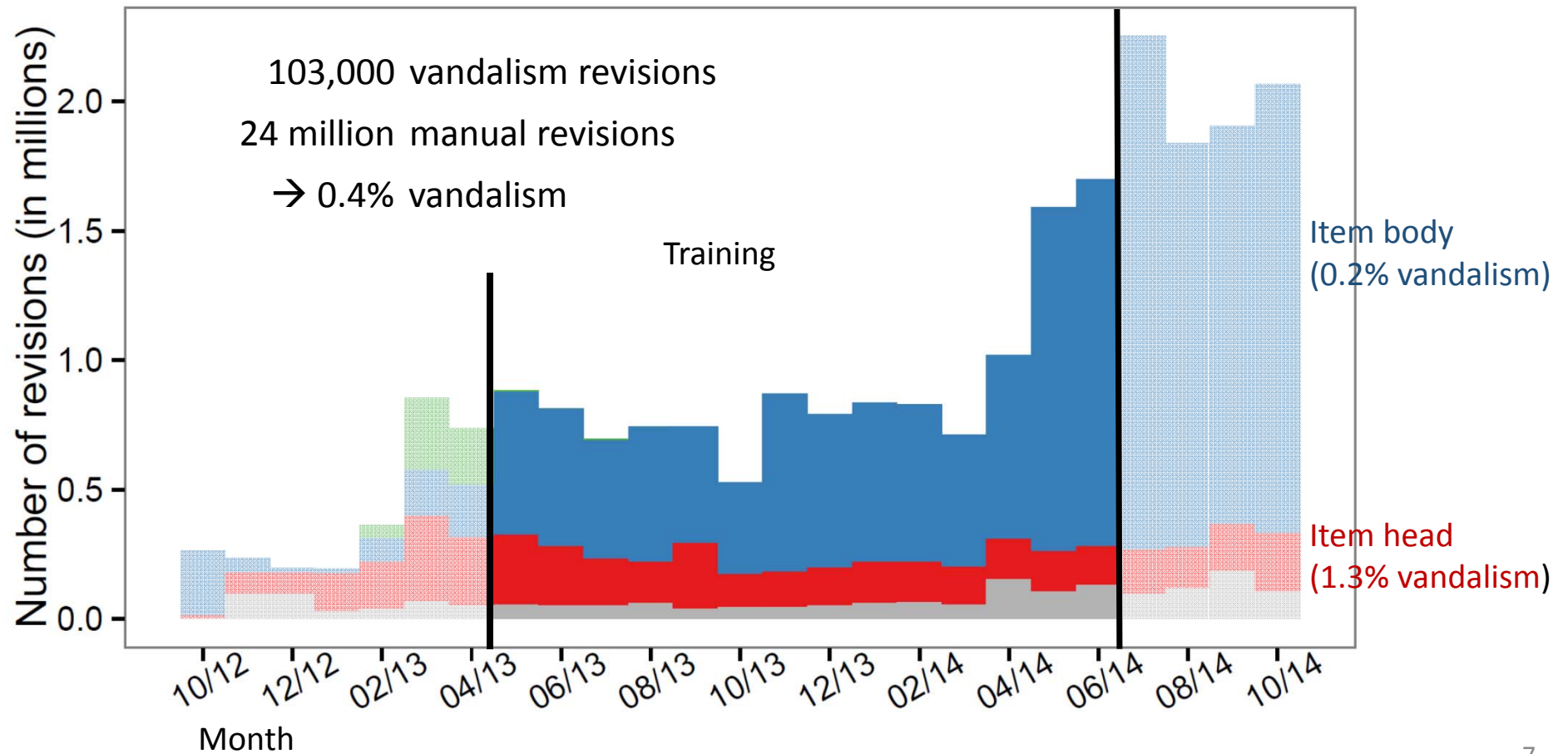
## Revisions over time



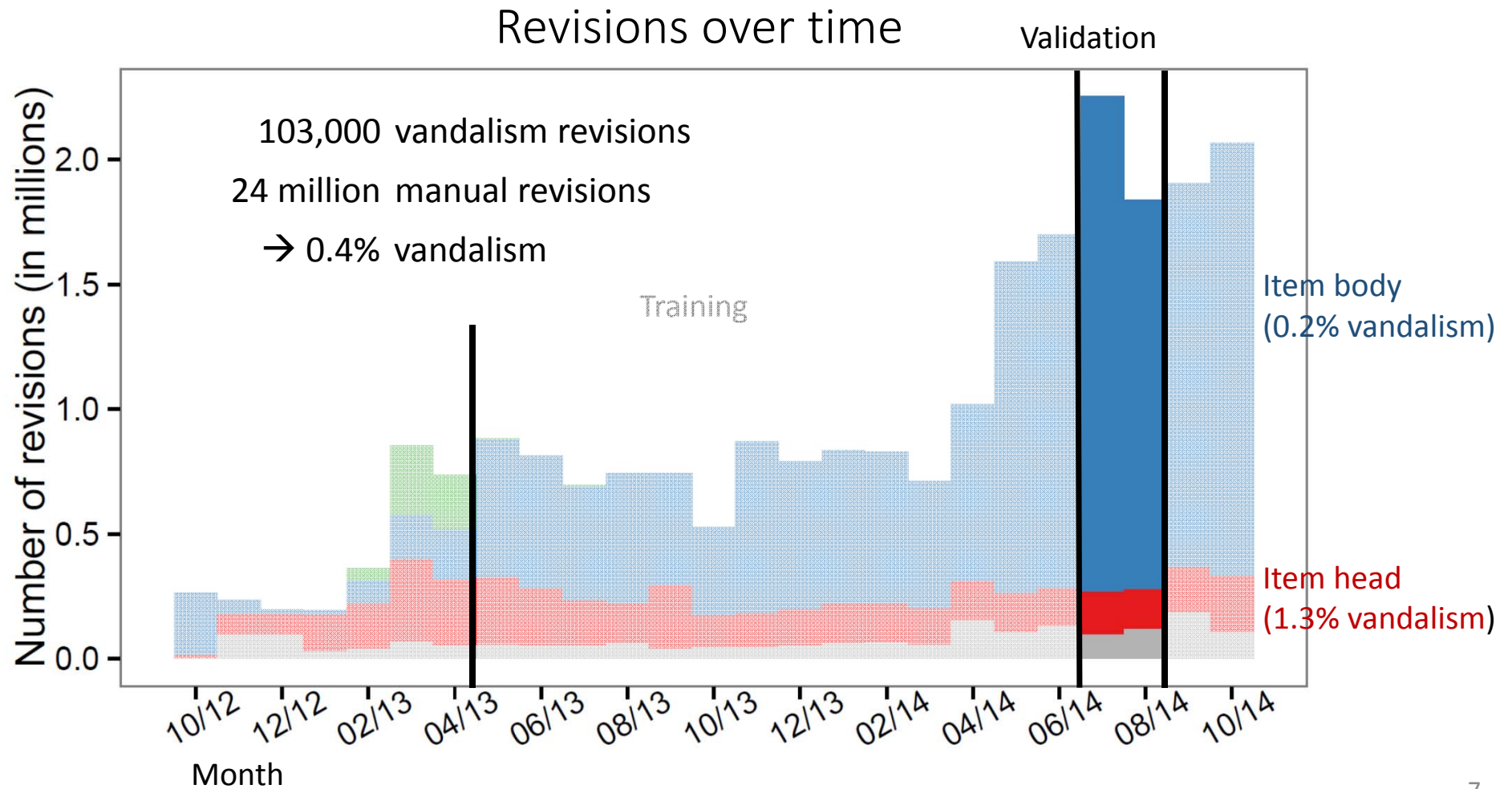


# Corpus [SIGIR '15]

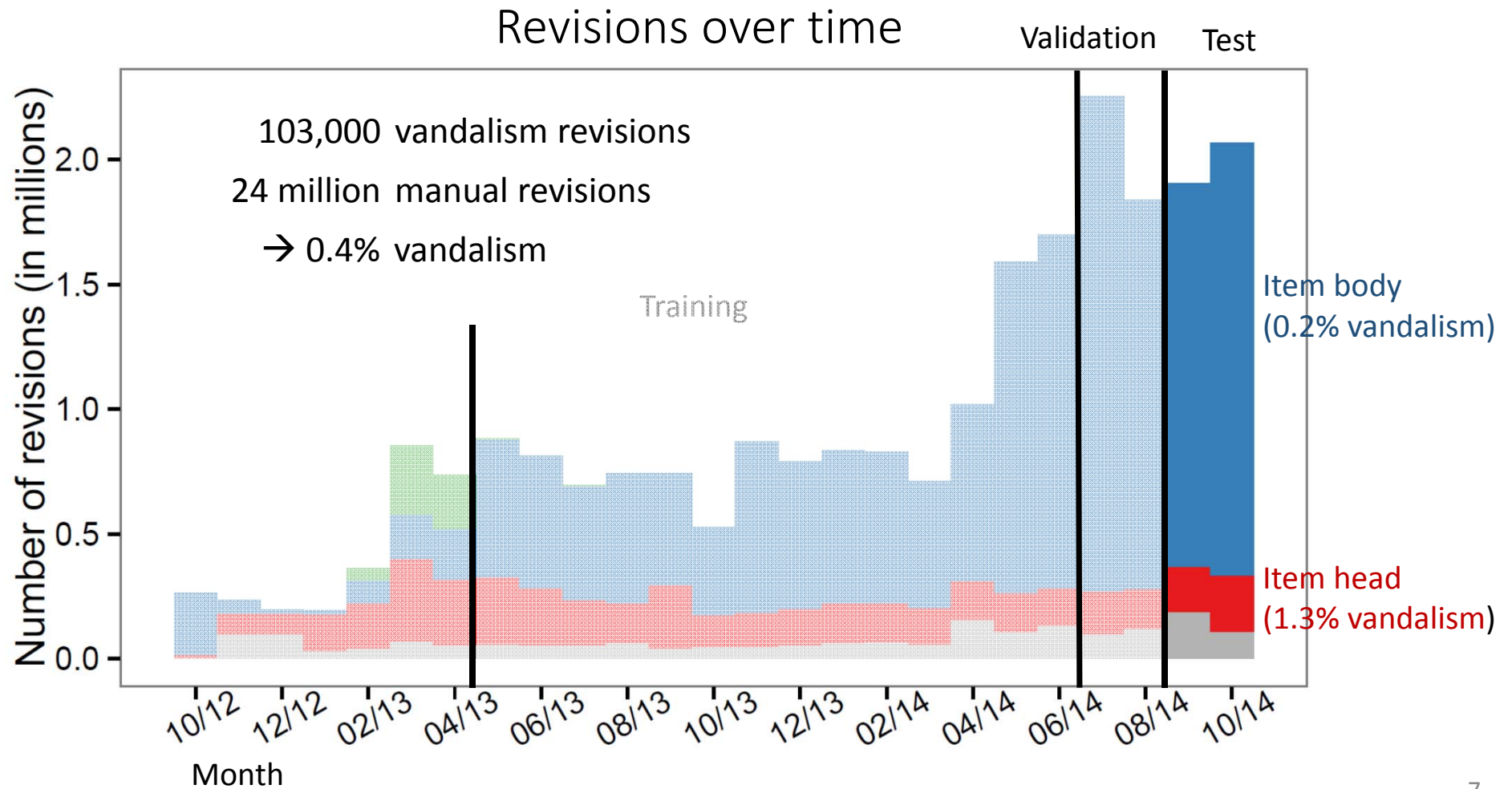
## Revisions over time



# Corpus [SIGIR '15]



# Corpus [SIGIR '15]



# Features (47 in total)

## Content Features

- 11 Character features (e.g., `lowerCaseRatio`, `digitRatio`)
- 9 Word features (e.g., `badWordRatio`)
- 4 Sentence features (e.g., `commentSitelinkSimilarity`)
- 3 Statement features (e.g., `propertyFrequency`)

## Context Features

- 10 User features (e.g., `userCountry`)
- 2 Item features (e.g., `logItemFrequency`)
- 8 Revision features (e.g., `revisionTag`, `revisionLanguage`)

# Features (47 in total)

## Content Features

- 11 Character features (e.g., `lowerCaseRatio`, `digitRatio`)
- 9 Word features (e.g., `badWordRatio`)
- 4 Sentence features (e.g., `commentSitelinkSimilarity`)
- 3 Statement features (e.g., `propertyFrequency`)

## Context Features

- 10 User features (e.g., `userCountry`)
- 2 Item features (e.g., `logItemFrequency`)
- 8 Revision features (e.g., **`revisionTag`**, `revisionLanguage`)

# Features (47 in total)

## Content Features

- 11 Character features
- 9 Word features
- 4 Sentence features
- 3 Statement features

<b>revisionTag</b>	<b>Vand.</b>	<b>Total</b>	<b>Prob.</b>
Rev. with tags	52 T	8,619 T	0.60%
Rev. w/o tags	52 T	15,386 T	0.34%

## Context Features

- 10 User features

- 2 Item features (e.g., `logItemFrequency`)
- 8 Revision features (e.g., **`revisionTag`**, `revisionLanguage`)

# Features (47 in total)

## Content Features

	<b>revisionTag</b>	<b>Vand.</b>	<b>Total</b>	<b>Prob.</b>
11 Character				
9 Word feat	Rev. with tags	52 T	8,619 T	0.60%
4 Sentence	<b>By abuse filter</b>	<b>49 T</b>	<b>122 T</b>	<b>39.90%</b>
3 Statement				

## Context Features

10 User feat	Rev. w/o tags	52 T	15,386 T	0.34%
--------------	---------------	------	----------	-------

2 Item features (e.g., `logItemFrequency`)

8 Revision features (e.g., **`revisionTag`**, `revisionLanguage`)

# Features (47 in total)

## Content Features

	<b>revisionTag</b>	<b>Vand.</b>	<b>Total</b>	<b>Prob.</b>
11 Character				
9 Word feat	Rev. with tags	52 T	8,619 T	0.60%
4 Sentence	<b>By abuse filter</b>	<b>49 T</b>	<b>122 T</b>	<b>39.90%</b>
3 Statement	<b>By editing tools</b>	<b>3 T</b>	<b>8,496 T</b>	<b>0.03%</b>
	Rev. w/o tags	52 T	15,386 T	0.34%

## Context Features

- 10 User features
- 2 Item features (e.g., `logItemFrequency`)
- 8 Revision features (e.g., **`revisionTag`**, `revisionLanguage`)



# Features (47 in total)

Details in Paper

## Content Features

- 11 Character features
- 9 Word features
- 4 Sentence features
- 3 Statement features

revisionTag	Vand.	Total	Prob.
Rev. with tags	52 T	8,619 T	0.60%
<b>By abuse filter</b>	<b>49 T</b>	<b>122 T</b>	<b>39.90%</b>
<b>By editing tools</b>	<b>3 T</b>	<b>8,496 T</b>	<b>0.03%</b>
Rev. w/o tags	52 T	15,386 T	0.34%

## Context Features

- 10 User features
- 2 Item features
- 8 Revision features

(e.g., `logItemFrequency`)

(e.g., **revisionTag**, `revisionLanguage`)

# Multiple-Instance Learning

# Multiple-Instance Learning

- Observation: Vandalism seldom occurs in isolation

# Multiple-Instance Learning

- Observation: Vandalism seldom occurs in isolation

22:35, 11 September 2013 184.19.64.111 (talk) . . (*Changed English label: Barack Obama Aloha*)

22:35, 11 September 2013 184.19.64.111 (talk) . . (*Added English alias: Lulu:):):):):):)*)

12:05, 11 September 2013 MatmaBot (talk | contribs) . . (*Changed Polish description: imported*)

# Multiple-Instance Learning

- Observation: Vandalism seldom occurs in isolation

22:35, 11 September 2013 184.19.64.111 (talk) . . (Changed English label: Barack Obama Aloha)

22:35, 11 September 2013 184.19.64.111 (talk) . . (Added English alias: Lulu:):):):):):))

12:05, 11 September 2013 MatmaBot (talk | contribs) . . (Changed Polish description: imported

# Multiple-Instance Learning

- Observation: Vandalism seldom occurs in isolation

**Session 1**

22:35, 11 September 2013 184.19.64.111 (talk) . . (Changed English label: Barack Obama Aloha)

22:35, 11 September 2013 184.19.64.111 (talk) . . (Added English alias: Lulu:):):):):))

**Session 2**

12:05, 11 September 2013 MatmaBot (talk | contribs) . . (Changed Polish description: imported

# Multiple-Instance Learning

- Observation: Vandalism seldom occurs in isolation

**Session 1**

22:35, 11 September 2013 184.19.64.111 (talk) . . (Changed English label: Barack Obama Aloha)

22:35, 11 September 2013 184.19.64.111 (talk) . . (Added English alias: Lulu:):):):):))

**Session 2**

12:05, 11 September 2013 MatmaBot (talk | contribs) . . (Changed Polish description: imported

- Idea: Apply Multiple-Instance Learning

# Multiple-Instance Learning

- Observation: Vandalism seldom occurs in isolation

**Session 1**

22:35, 11 September 2013 184.19.64.111 (talk) . . (Changed English label: Barack Obama Aloha)

22:35, 11 September 2013 184.19.64.111 (talk) . . (Added English alias: Lulu:):):):):))

**Session 2**

12:05, 11 September 2013 MatmaBot (talk | contribs) . . (Changed Polish description: imported

- Idea: Apply Multiple-Instance Learning



# Multiple-Instance Learning

- Observation: Vandalism seldom occurs in isolation

Session 1

22:35, 11 September 2013 184.19.64.111 (talk) . . (Changed English label: Barack Obama Aloha)

22:35, 11 September 2013 184.19.64.111 (talk) . . (Added English alias: Lulu:):):):):))

Session 2

12:05, 11 September 2013 MatmaBot (talk | contribs) . . (Changed Polish description: imported

- Idea: Apply Multiple-Instance Learning

# WDVD vs. Baselines

- WDVD (our approach)  
**Wikidata Vandalism Detector**

# WDVD vs. Baselines

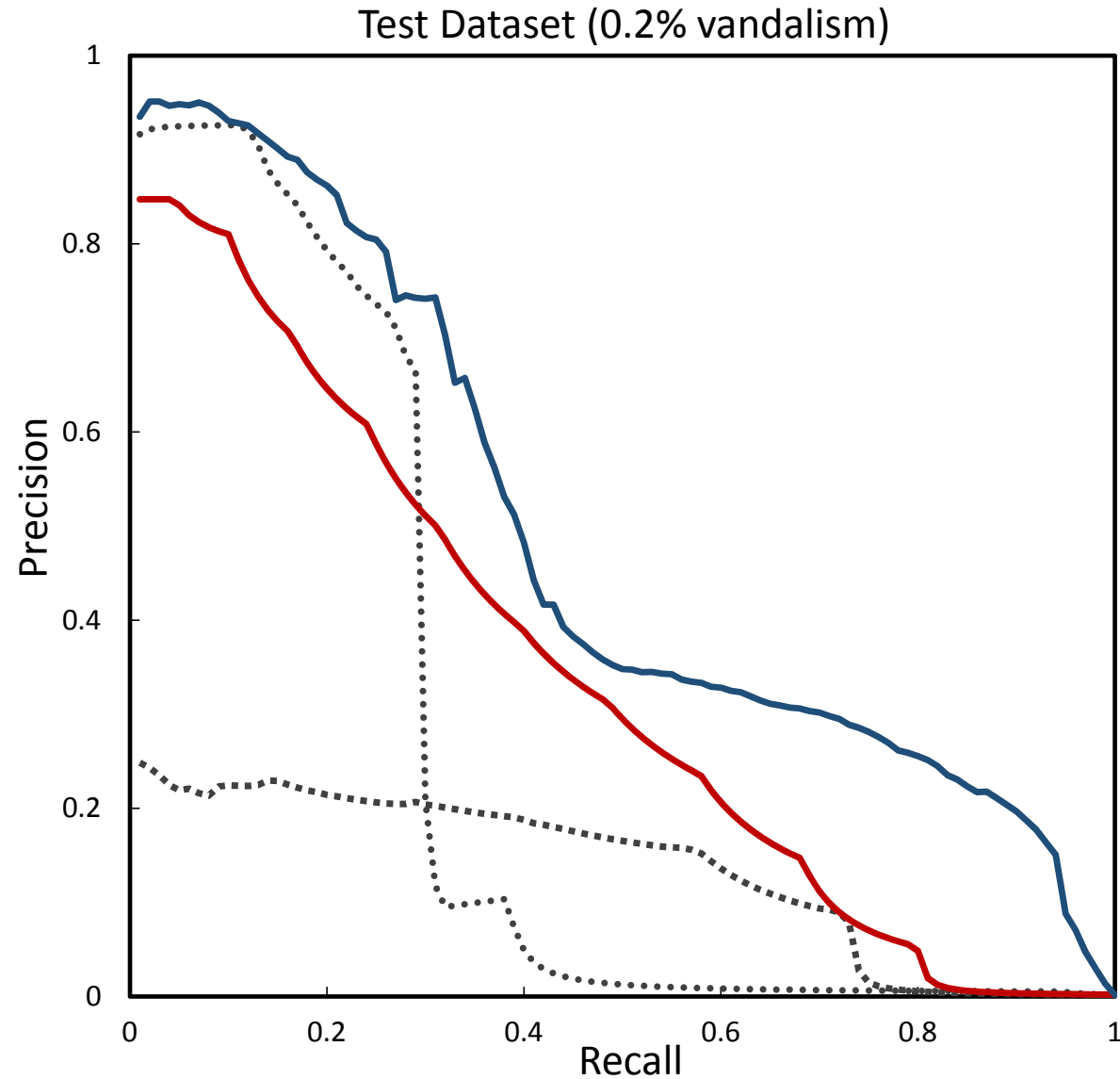
- **WDVD** (our approach)  
Wikidata **V**andalism **D**etector
- **FILTER** (baseline)  
Wikidata Abuse **F**ilter

# WDVD vs. Baselines

- **WDVD** (our approach)  
**Wikidata Vandalism Detector**
- **FILTER** (baseline)  
Wikidata Abuse **Filter**
- **ORES** (baseline)  
**Objective Revision Evaluation Service**

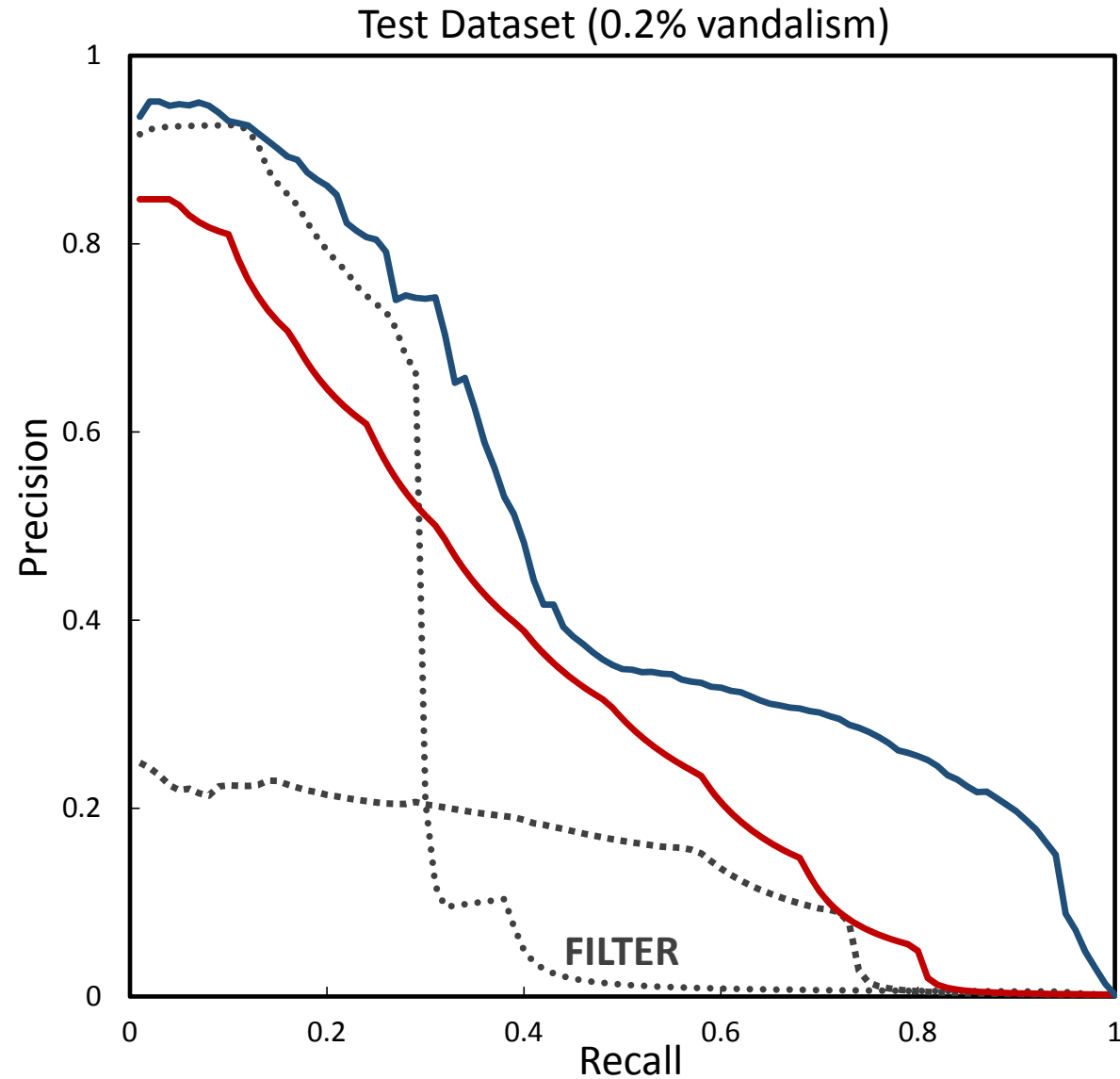
# WDVD vs. Baselines

- **WDVD (our approach)**  
**Wikidata Vandalism Detector**
- **FILTER (baseline)**  
**Wikidata Abuse Filter**
- **ORES (baseline)**  
**Objective Revision Evaluation Service**



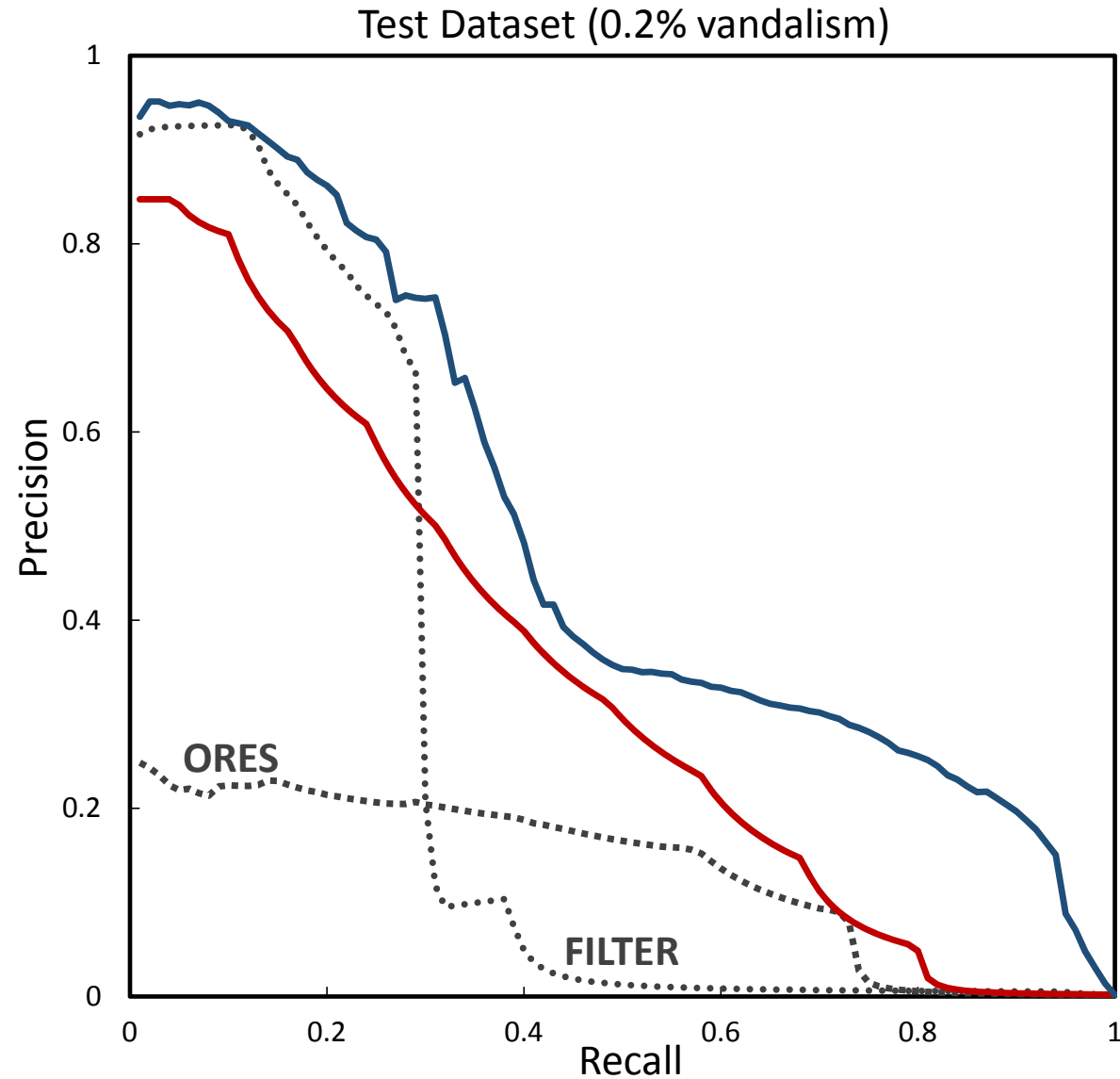
# WDVD vs. Baselines

- **WDVD (our approach)**  
**Wikidata Vandalism Detector**
- **FILTER (baseline)**  
**Wikidata Abuse Filter**
- **ORES (baseline)**  
**Objective Revision Evaluation Service**



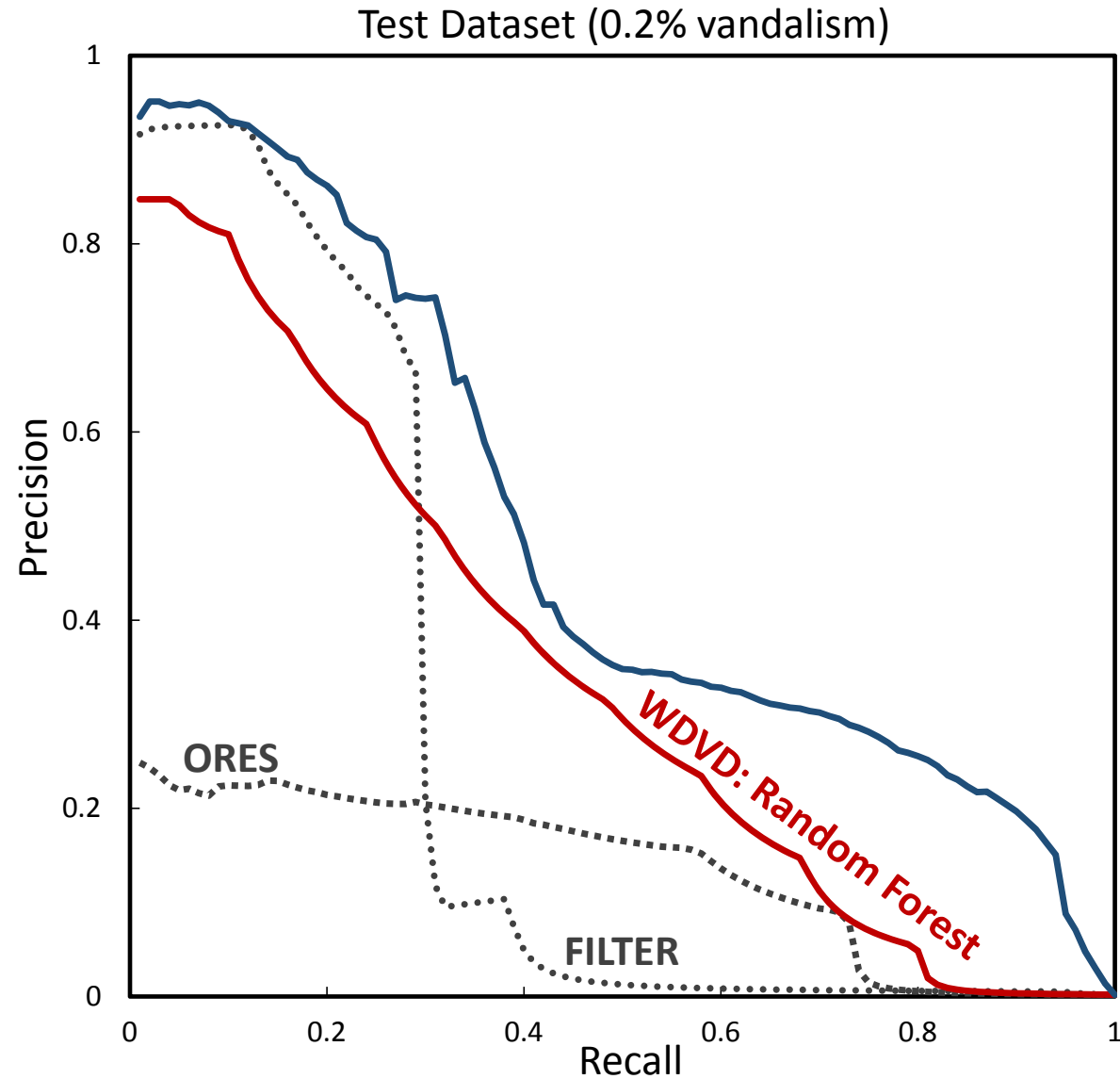
# WDVD vs. Baselines

- **WDVD (our approach)**  
**Wikidata Vandalism Detector**
- **FILTER (baseline)**  
**Wikidata Abuse Filter**
- **ORES (baseline)**  
**Objective Revision Evaluation Service**



# WDVD vs. Baselines

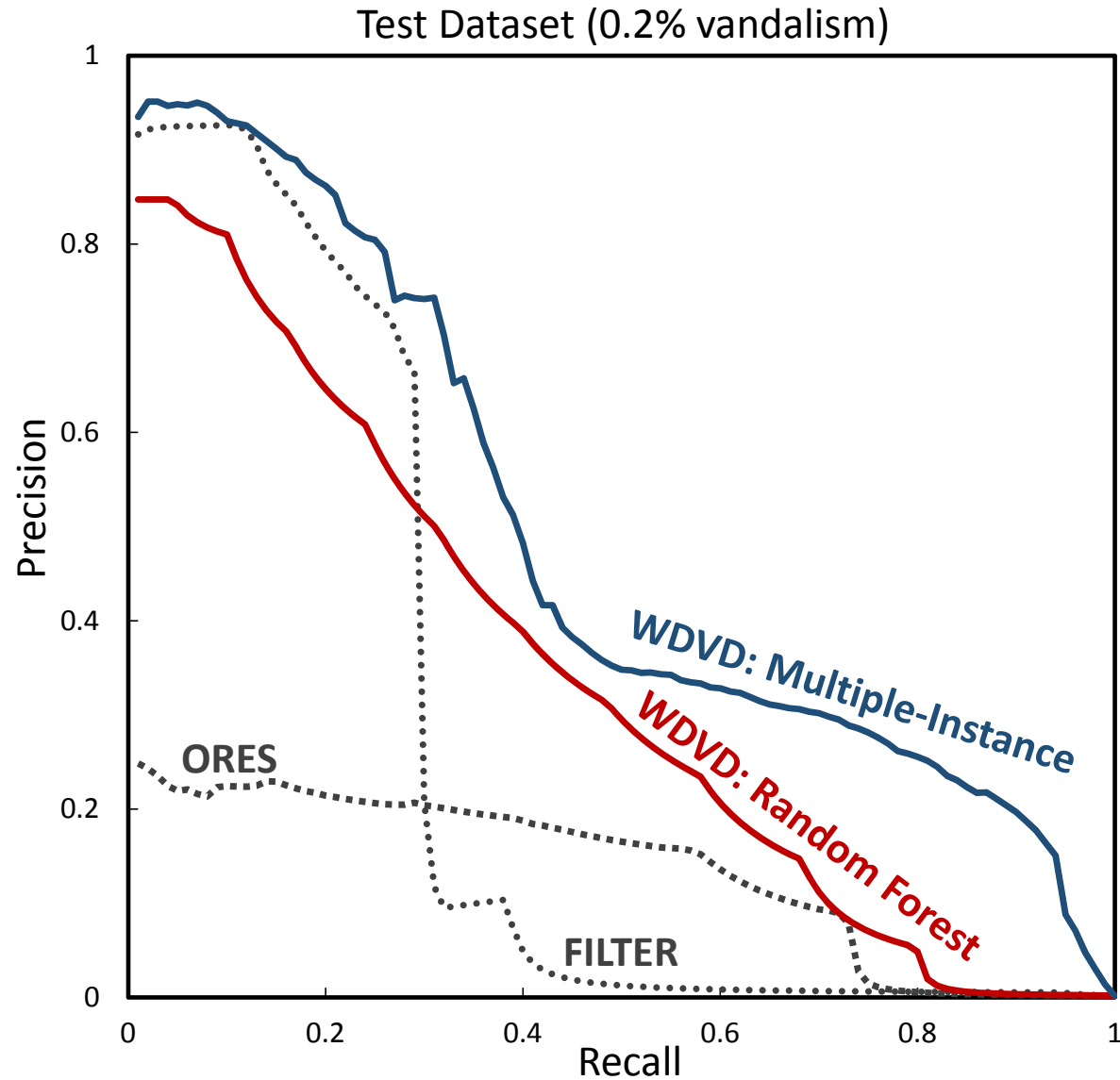
- **WDVD (our approach)**  
**Wikidata Vandalism Detector**
- **FILTER (baseline)**  
**Wikidata Abuse Filter**
- **ORES (baseline)**  
**Objective Revision Evaluation Service**





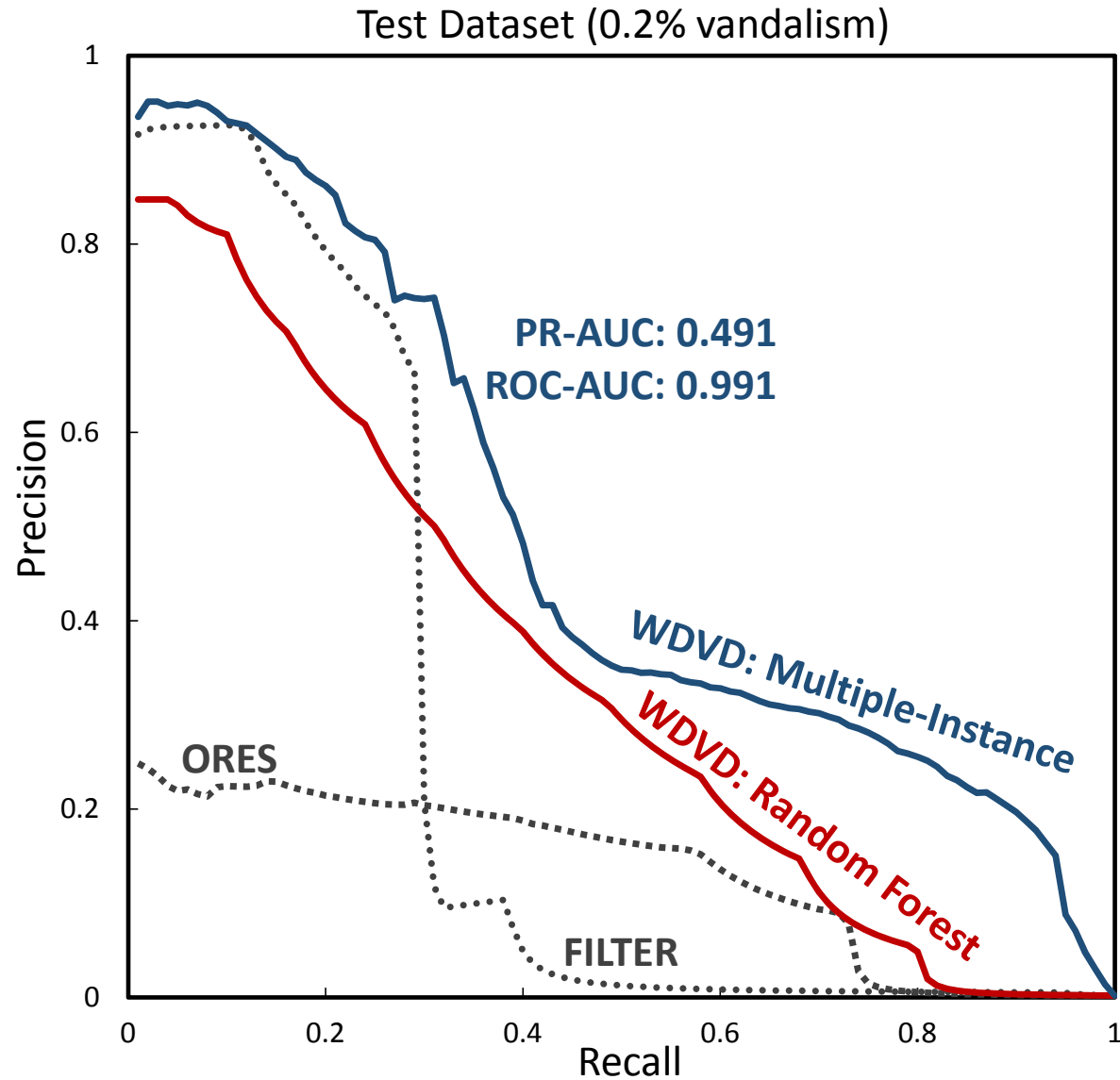
# WDVD vs. Baselines

- **WDVD (our approach)**  
**Wikidata Vandalism Detector**
- **FILTER (baseline)**  
**Wikidata Abuse Filter**
- **ORES (baseline)**  
**Objective Revision Evaluation Service**



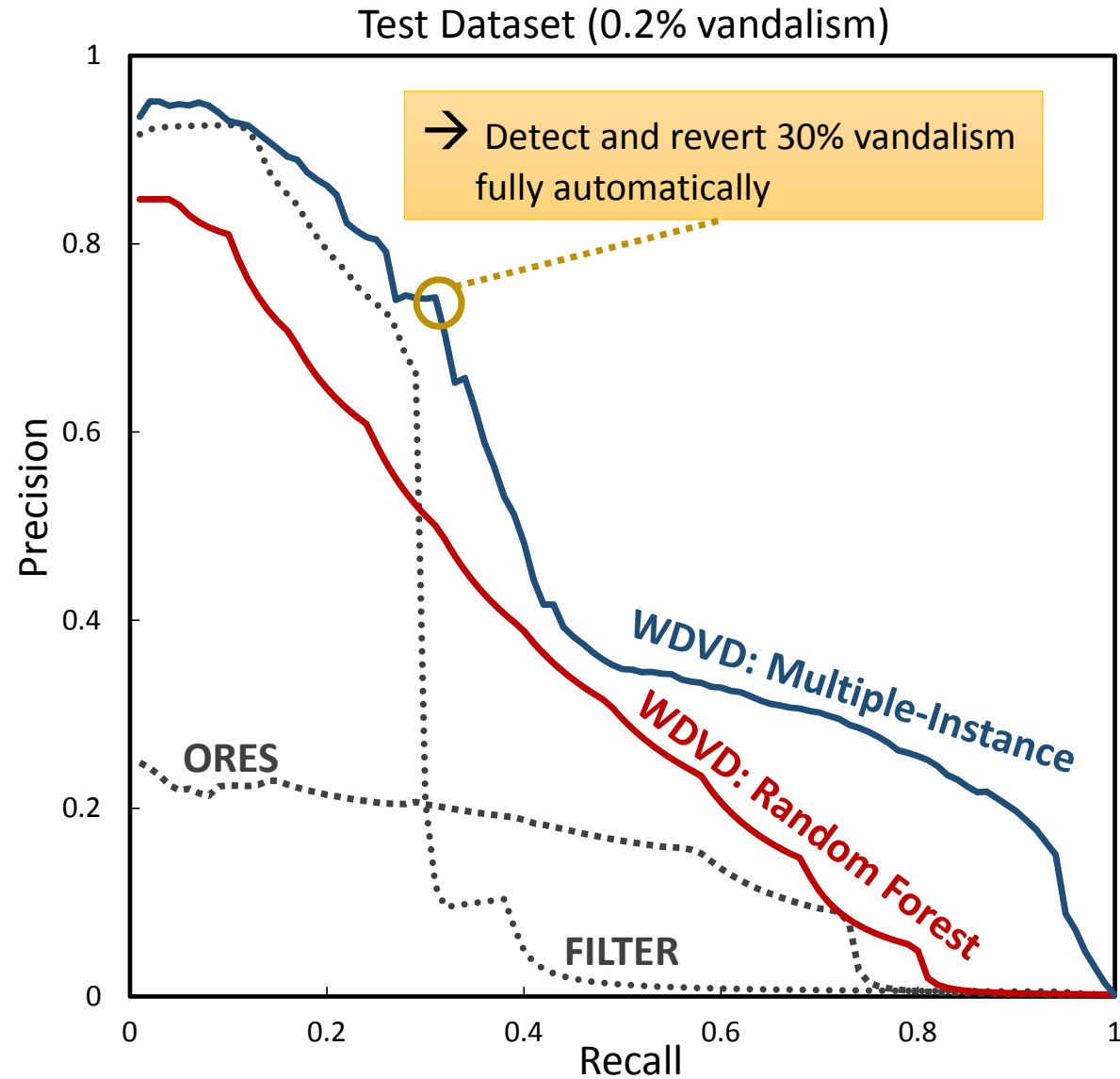
# WDVD vs. Baselines

- **WDVD (our approach)**  
**Wikidata Vandalism Detector**
- **FILTER (baseline)**  
**Wikidata Abuse Filter**
- **ORES (baseline)**  
**Objective Revision Evaluation Service**



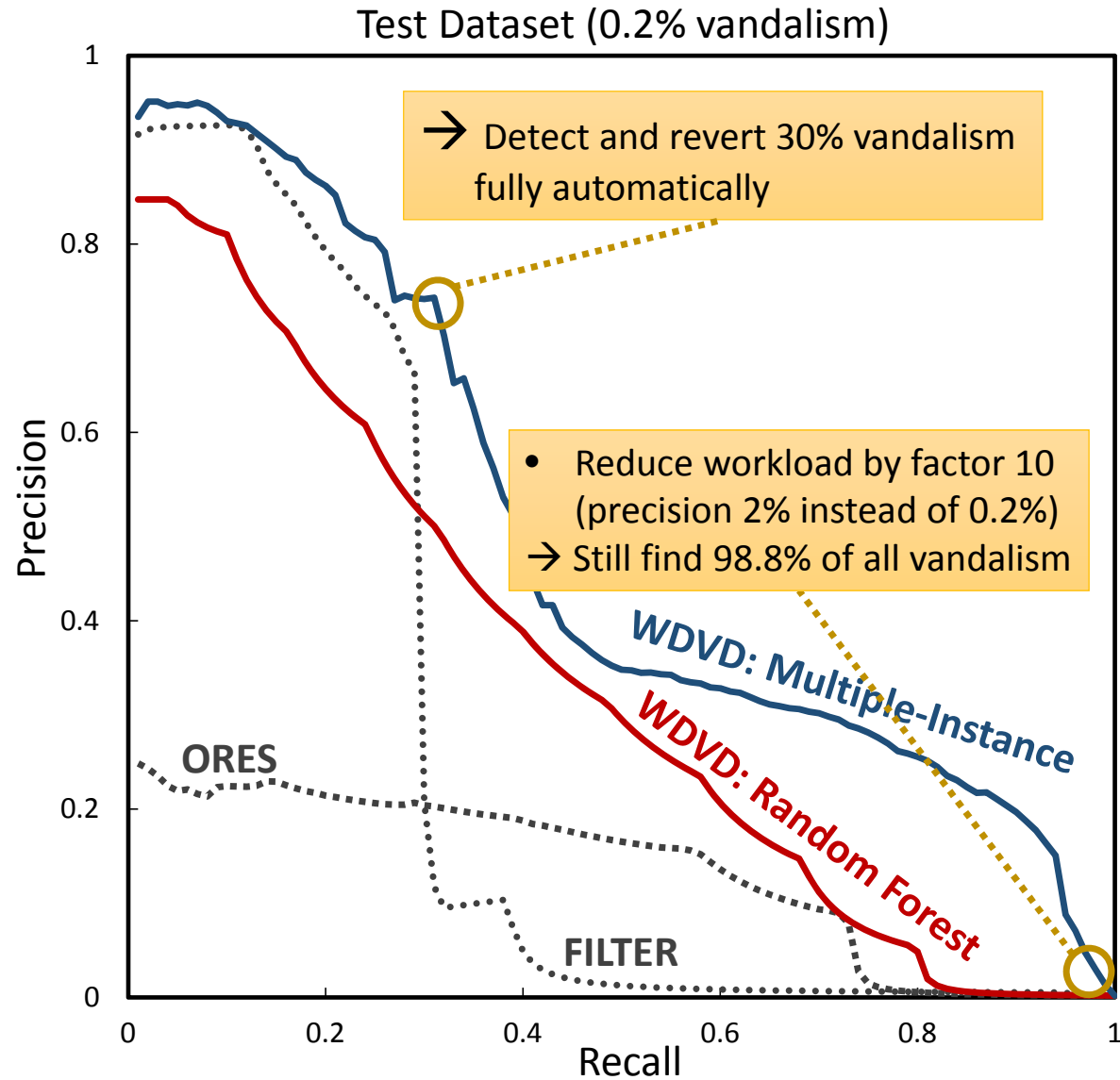
# WDVD vs. Baselines

- **WDVD (our approach)**  
**W**ikidata **V**andalism **D**etector
- **FILTER (baseline)**  
Wikidata Abuse **F**ilter
- **ORES (baseline)**  
**O**bjective **R**evision **E**valuation **S**ervice



# WDVD vs. Baselines

- **WDVD (our approach)**  
**Wikidata Vandalism Detector**
- **FILTER (baseline)**  
**Wikidata Abuse Filter**
- **ORES (baseline)**  
**Objective Revision Evaluation Service**



# Conclusion and Outlook

# Conclusion and Outlook

## Conclusion

- Vandalism: Concentration on item heads (currently)
- Features: Content & Context
- Model: Multiple-Instance
- PR-AUC: 0.491
- ROC-AUC: 0.991

# Conclusion and Outlook



Code + Data:

<http://www.heindorf.me/wdvd.html>

## Conclusion

- Vandalism: Concentration on item heads (currently)
- Features: Content & Context
- Model: Multiple-Instance
- PR-AUC: 0.491
- ROC-AUC: 0.991

# Conclusion and Outlook



Code + Data:

<http://www.heindorf.me/wdvd.html>

## Conclusion

- Vandalism: Concentration on item heads (currently)
- Features: Content & Context
- Model: Multiple-Instance
- PR-AUC: 0.491
- ROC-AUC: 0.991

## Outlook

- Goal: Better detection (on item bodies)
- Idea: Double-check with other sources



# Conclusion and Outlook



Code + Data:

<http://www.heindorf.me/wdvd.html>

## Conclusion

- Vandalism: Concentration on item heads (currently)
- Features: Content & Context
- Model: Multiple-Instance
- PR-AUC: 0.491
- ROC-AUC: 0.991



Join the competition:

Vandalism Detection @WSDM Cup 2017

<http://www.wsdm-cup-2017.org/>

## Outlook

- Goal: Better detection (on item bodies)
- Idea: Double-check with other sources

# Conclusion and Outlook



Code + Data:

<http://www.heindorf.me/wdvd.html>

## Conclusion

- Vandalism: Concentration on item heads (currently)
- Features: Content & Context
- Model: Multiple-Instance
- PR-AUC: 0.491
- ROC-AUC: 0.991



Join the competition:

Vandalism Detection @WSDM Cup 2017

<http://www.wsdm-cup-2017.org/>

## Outlook

- Goal: Better detection (on item bodies)
- Idea: Double-check with other sources

## Acknowledgement

- German Research Foundation (DFG)
- SIGIR Student Travel Grant

# Conclusion and Outlook



Code + Data:  
<http://www.heindorf.me/wdvd.html>

## Conclusion

- Vandalism: Concentration on item heads (currently)
- Features: Content & Context
- Model: Multiple-Instance
- PR-AUC: 0.491
- ROC-AUC: 0.991



Join the competition:  
Vandalism Detection @WSDM Cup 2017  
<http://www.wsdm-cup-2017.org/>

## Outlook

- Goal: Better detection (on item bodies)
- Idea: Double-check with other sources

## Acknowledgement

- German Research Foundation (DFG)
- SIGIR Student Travel Grant

# Thank you!