

# Overview of the 1st International Competition on Quality Flaw Prediction in Wikipedia

---

Maik Anderka      Benno Stein

Bauhaus-Universität Weimar  
<http://pan.webis.de>



# Wikipedia Facts

- 285 languages
- 87 339 125 pages
- 23 013 694 encyclopedic articles
- 2 029 274 images
- 1 416 124 240 edits
- 36 279 901 registered users
- 4 554 admins



Launched in January 2001,  
`wikipedia.org` is the sixth most-visited website.

[<http://www.alexa.com/siteinfo/wikipedia.org>]

[[http://meta.wikimedia.org/wiki/List\\_of\\_Wikipedias](http://meta.wikimedia.org/wiki/List_of_Wikipedias)]

# What about Information Quality?

# What about Information Quality?

- ❑ Everyone can edit Wikipedia, even anonymously
  - ❑ Heterogeneous community of Wikipedia authors
  - ❑ Edits are not reviewed before publication
- Extremely varying content quality

Two key objectives:

1. Improve low-quality content
2. Maintain high-quality content

# Automatic Quality Assessment

- Up to now: classification into abstract quality schemes
- For instance “Is an article featured or not?”

[Hu et al., CIKM 2007]

[Wilkinson and Huberman, WikiSym 2007]

[Blumenstock, WWW 2008]

[Dalip et al., JCDL 2009]

[Likpa and Stein, WWW 2010]

- Classifiers perform nearly perfect, but
  - No rationale why an article violates Wikipedia’s featured article criteria
  - No practical support for Wikipedia’s quality assurance process

# Automatic Quality Assessment

- Up to now: classification into abstract quality schemes
- For instance “Is an article featured or not?”

[Hu et al., CIKM 2007]

[Wilkinson and Huberman, WikiSym 2007]

[Blumenstock, WWW 2008]

[Dalip et al., JCDL 2009]

[Likpa and Stein, WWW 2010]

- Classifiers perform nearly perfect, but
  - No rationale why an article violates Wikipedia’s featured article criteria
  - No practical support for Wikipedia’s quality assurance process

Less than 0.1% of the English Wikipedia articles are featured

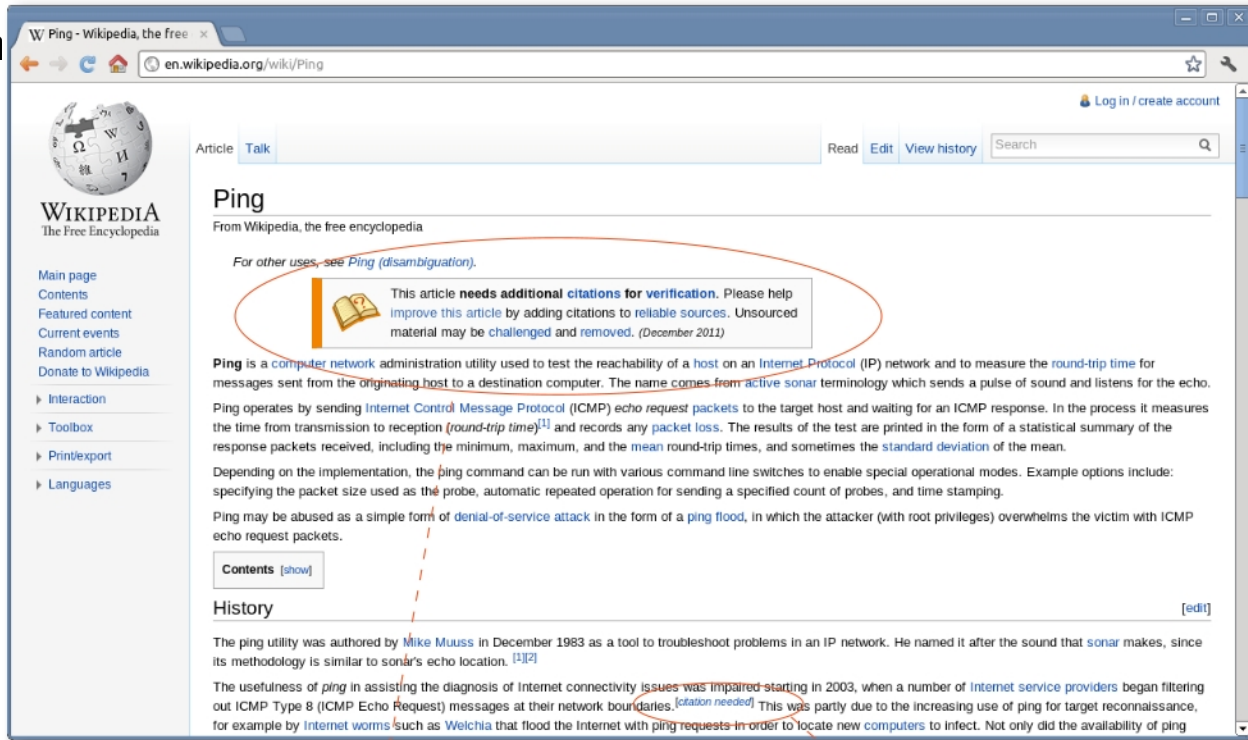
What is wrong with the remaining 99.9%?

# Previous Work

Use cleanup tags to analyze quality flaws. [Anderka et al., WWW 2011]

# Previous Work

Use clean



The screenshot shows the Wikipedia article for "Ping". A red circle highlights a notice box that reads: "This article needs additional citations for verification. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (December 2011)". The article text includes: "Ping is a computer network administration utility used to test the reachability of a host on an Internet Protocol (IP) network and to measure the round-trip time for messages sent from the originating host to a destination computer. The name comes from active sonar terminology which sends a pulse of sound and listens for the echo. Ping operates by sending Internet Control Message Protocol (ICMP) echo request packets to the target host and waiting for an ICMP response. In the process it measures the time from transmission to reception (round-trip time)<sup>[1]</sup> and records any packet loss. The results of the test are printed in the form of a statistical summary of the response packets received, including the minimum, maximum, and the mean round-trip times, and sometimes the standard deviation of the mean. Depending on the implementation, the ping command can be run with various command line switches to enable special operational modes. Example options include: specifying the packet size used as the probe, automatic repeated operation for sending a specified count of probes, and time stamping. Ping may be abused as a simple form of denial-of-service attack in the form of a ping flood, in which the attacker (with root privileges) overwhelms the victim with ICMP echo request packets." Below the article is a "History" section with a "[edit]" link.

[citation needed]



This article **does not cite any references or sources**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (December 2011)



# Previous Work

Use cleanup tags to analyze quality flaws. [Anderka et al., WWW 2011]

- Exploratory analysis of the English Wikipedia:
  - 388 cleanup tags
  - 27.53% of all articles are tagged with at least one flaw
  - 70% of the tagged flaws concern verifiability of information
  - The actual number of flaws is even higher

# Previous Work

Use cleanup tags to analyze quality flaws. [Anderka et al., WWW 2011]

- Exploratory analysis of the English Wikipedia:
  - 388 cleanup tags
  - 27.53% of all articles are tagged with at least one flaw
  - 70% of the tagged flaws concern verifiability of information
  - The actual number of flaws is even higher

But, how to predict quality flaws of untagged articles?

# Task Description

# Task Description

## Problem Statement

“Decide whether or not an article contains a quality flaw  $f$ , given a sample of articles containing  $f$ .”

Key challenges:

- ❑ Only positive examples are available (articles tagged with flaw  $f$ )
- ❑ A co-class cannot be modeled
- ❑ No *representative* sample of articles not containing  $f$

# Task Description

## Problem Statement

“Decide whether or not an article contains a quality flaw  $f$ , given a sample of articles containing  $f$ .”

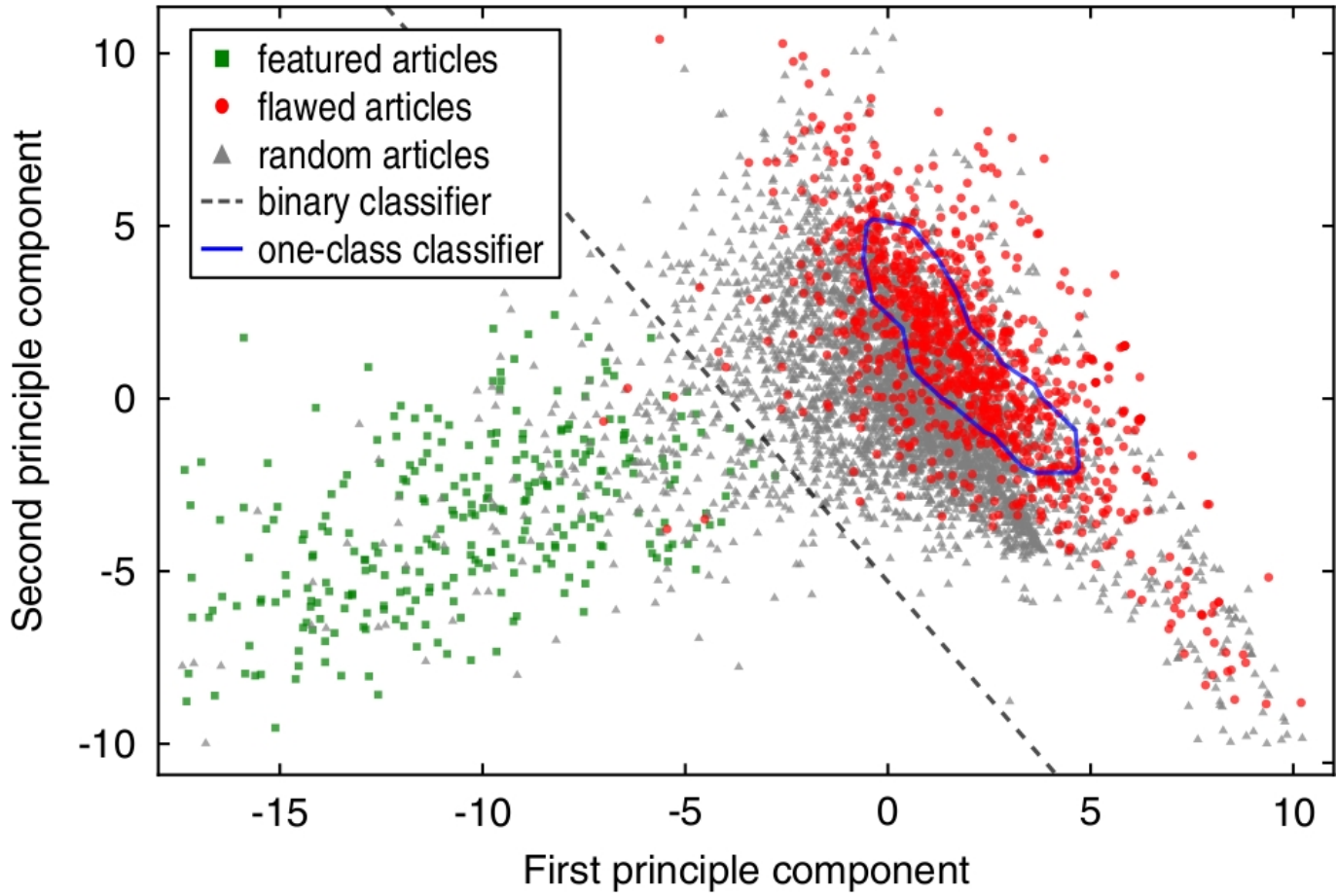
Key challenges:

- ❑ Only positive examples are available (articles tagged with flaw  $f$ )
- ❑ A co-class cannot be modeled
- ❑ No *representative* sample of articles not containing  $f$

→ One-class problem

# Task Description

## Problem Statement



# Task Description

## Quality Flaws

- The task targets ten important quality flaws of English Wikipedia articles
- The prediction performance is evaluated individually for each flaw

---

<b>Flaw name</b>	<b>Description</b>
Unreferenced	The article does not cite any references or sources.
Orphan	The article has fewer than three incoming links.
Refimprove	The article needs additional citations for verification.
Empty section	The article has at least one section that is empty.
Notability	The article does not meet the general notability guideline.
No footnotes	The article's sources remain unclear because of its inline citations.
Primary sources	The article relies on references to primary sources.
Wikify	The article needs to be wikified (internal links and layout).
Advert	The article is written like an advertisement.
Original research	The article contains original research.

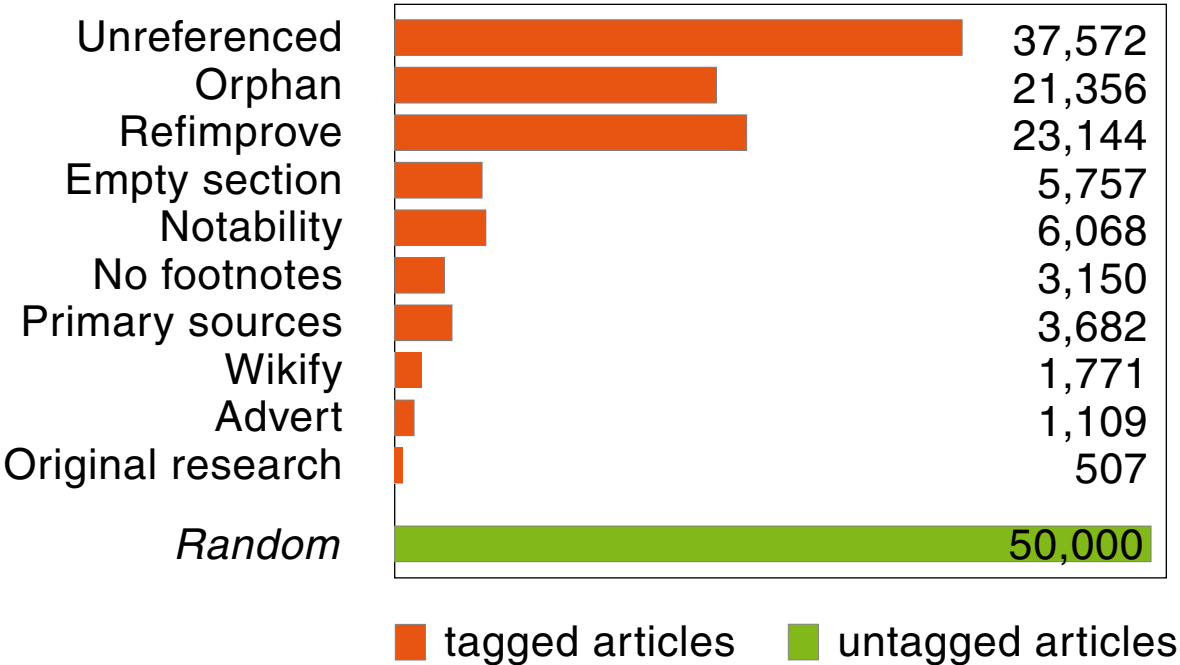
---

# Task Description

## Data

173,126 English Wikipedia articles (snapshot from January 4th, 2012)

### Training corpus

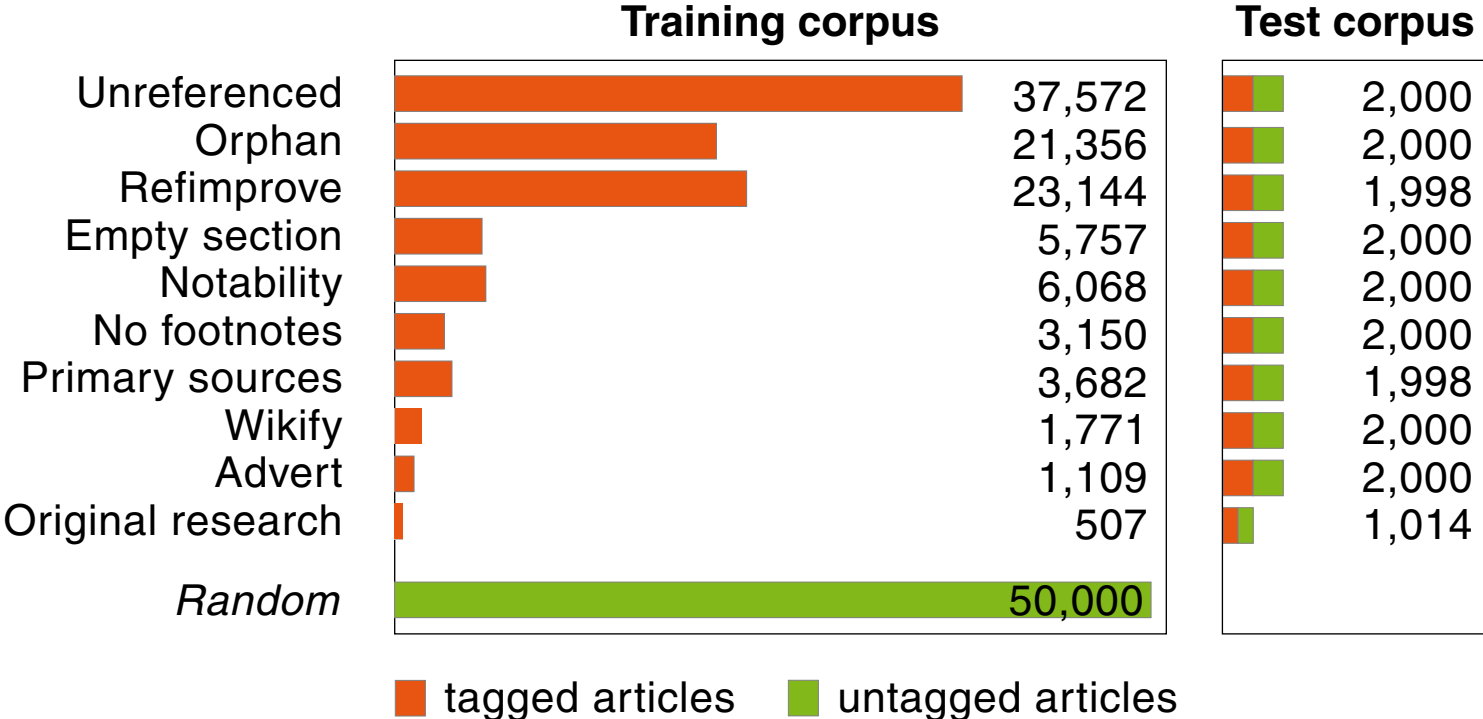




# Task Description

## Data

173,126 English Wikipedia articles (snapshot from January 4th, 2012)



# Results

# Results

## Participants

- 21 registered teams
- 3 teams submitted runs

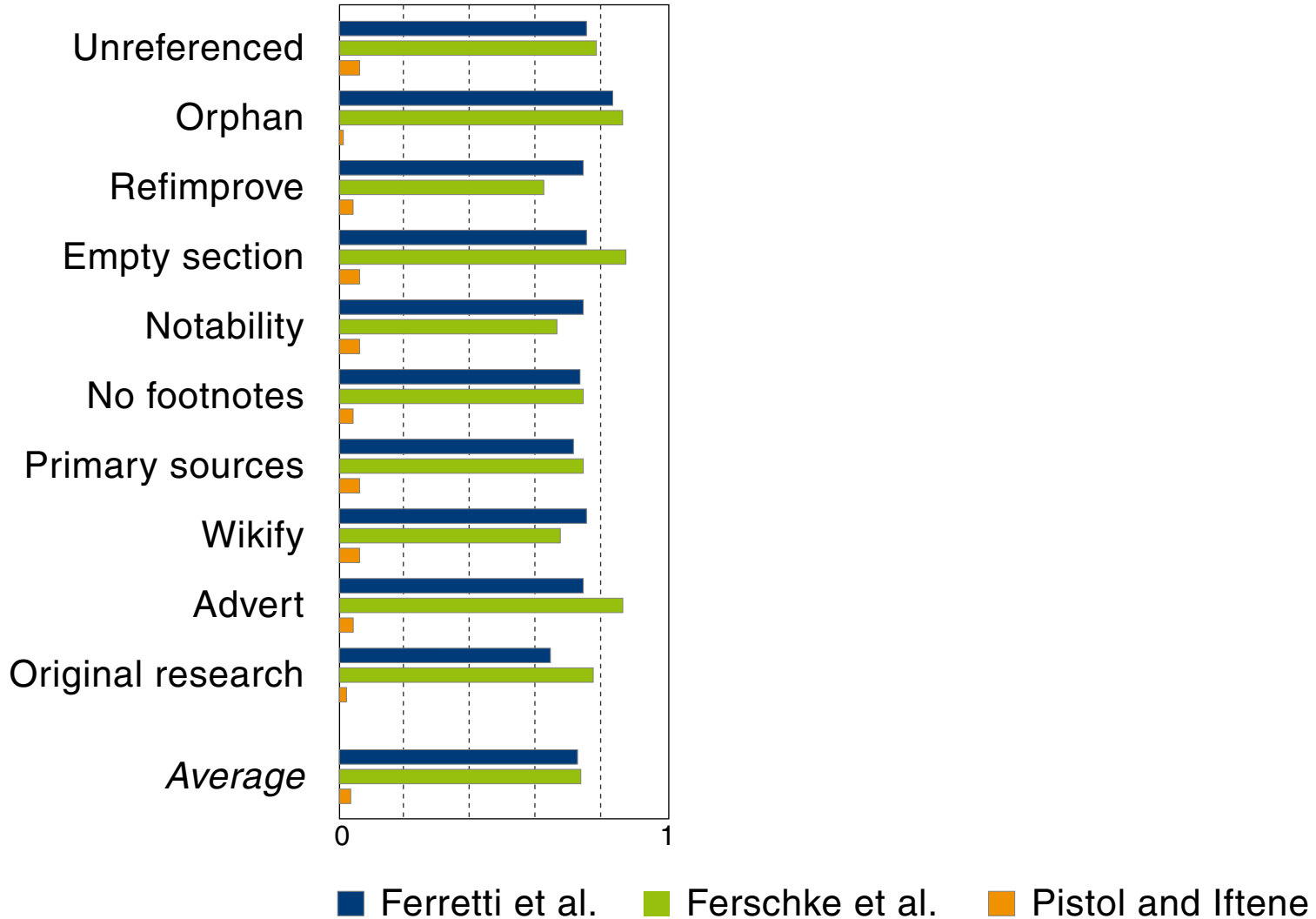
---

<b>Team name</b>	<b>Participants and affiliations</b>
Ferretti et al.	Edgardo Ferretti <sup>*</sup> , Donato Hernández Fusilier <sup>°</sup> , Rafael Guzmán Cabrera <sup>°</sup> , Manuel Montes-y-Gómez <sup>†</sup> , Marcelo Errecalde <sup>*</sup> , and Paolo Rosso <sup>‡</sup> <sup>*</sup> Universidad Nacional de San Luis, Argentina <sup>°</sup> Universidad de Guanajuato, Mexico <sup>†</sup> Óptica y Electrónica (INAOE), Mexico <sup>‡</sup> Universidad Politécnica de Valencia, Spain
Ferschke et al.	Oliver Ferschke, Iryna Gurevych, and Marc Rittberger Technische Universität Darmstadt, Germany
Pistol and Iftene	Ionut Cristian Pistol and Adrian Iftene “Alexandru Ioan Cuza” University of Iasi, Romania

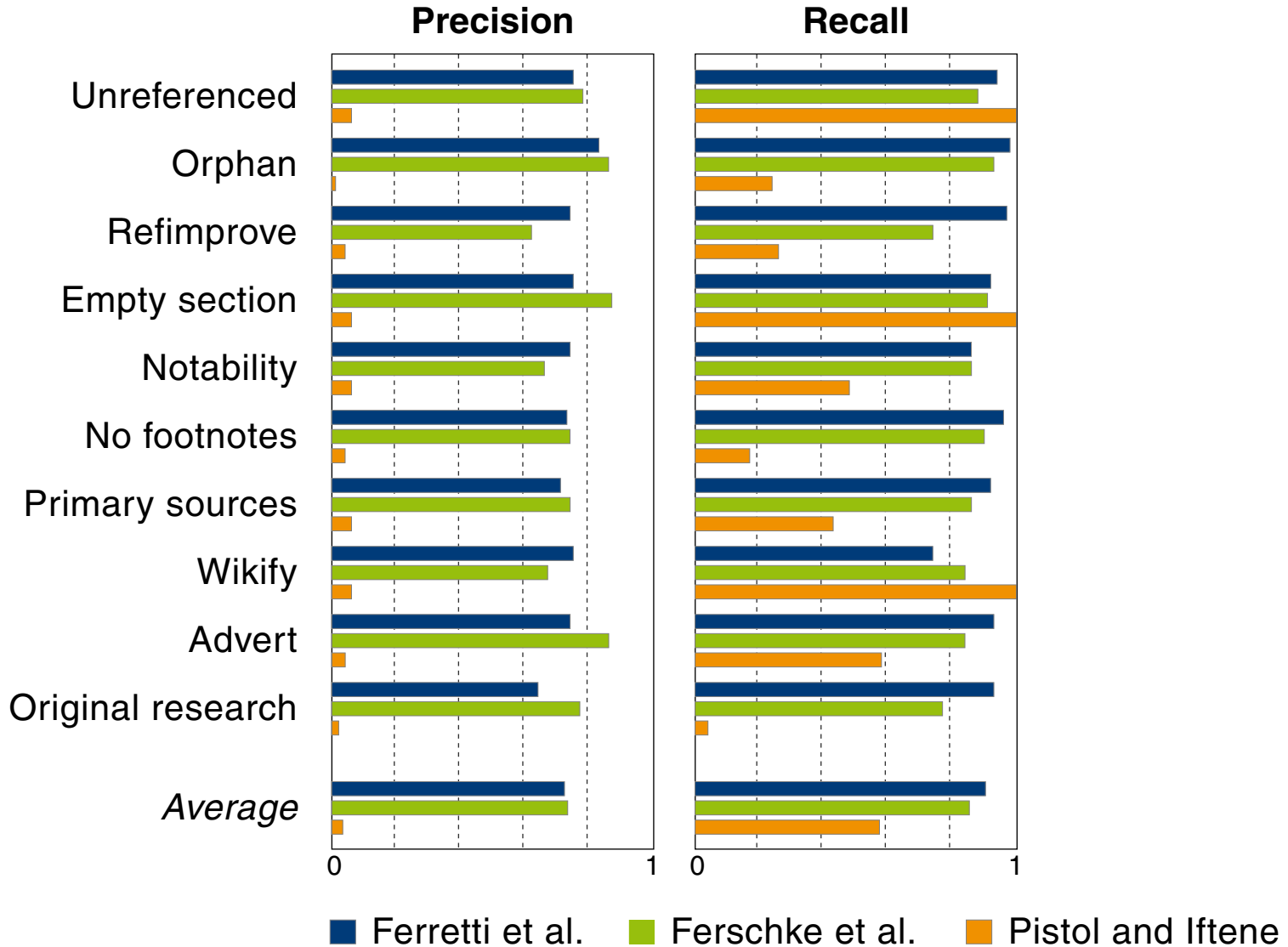
---

# Results

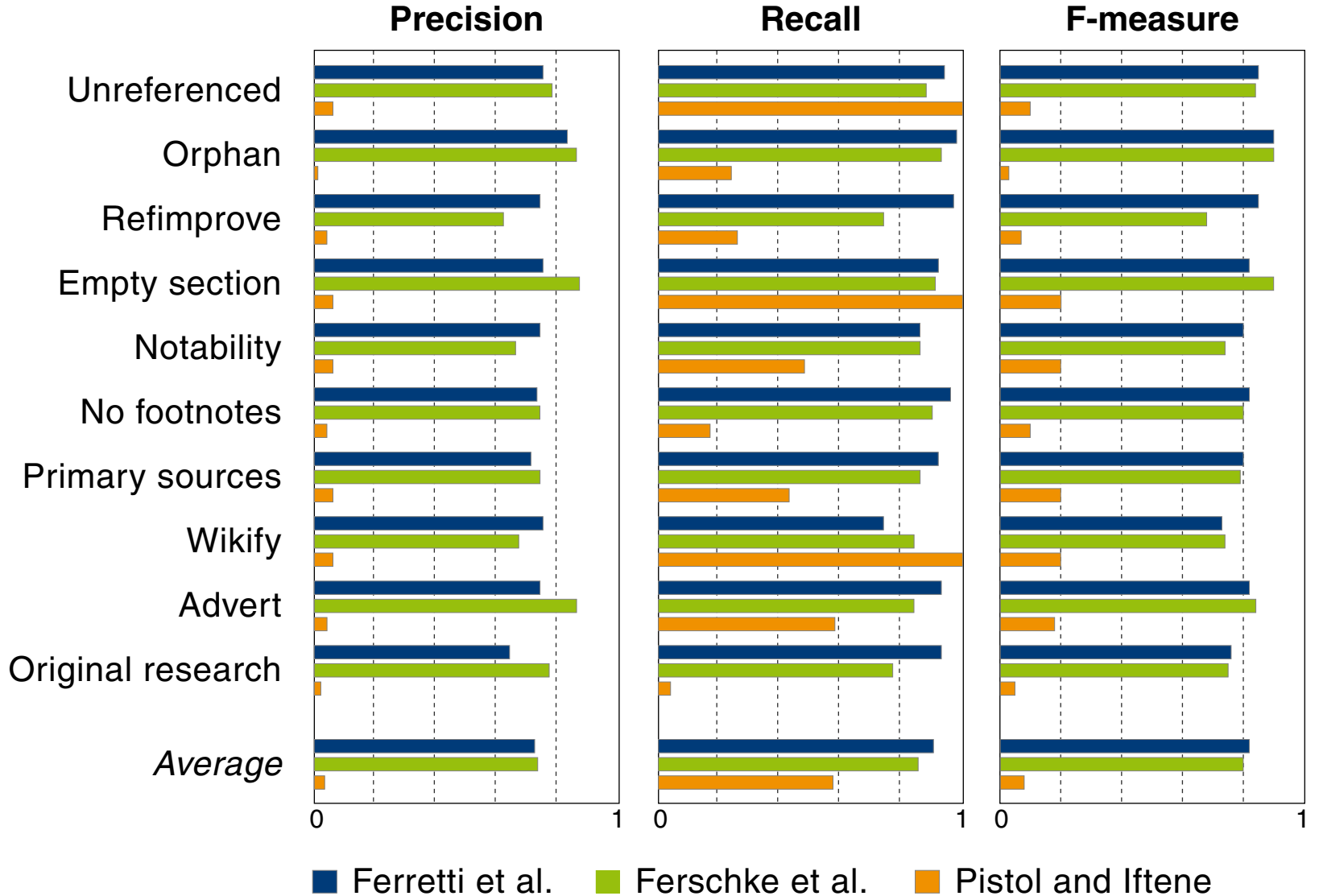
## Precision



# Results



# Results



# Conclusion

- What we got
  - Three quality flaw classifiers from which two achieve a promising effectiveness for particular flaws
  - First corpus of flawed Wikipedia articles: PAN Wikipedia quality flaw corpus 2012 (PAN-WQF-12)
  
- Lessons learned
  - This task subsumes the vandalism detection task of previous years
  - Promising performance for particular flaws
  - More flaw types need to be investigated
  - Automatic tagging of quality flaws in Wikipedia within reach

# 1st International Competition on Quality Flaw Prediction in Wikipedia

---

## Winner:

Edgardo Ferretti, Donato Hernández Fusilier,  
Rafael Guzmán Cabrera, Manuel Montes-y-Gómez,  
Marcelo Errecalde, and Paolo Rosso

