

Citation Needed: A Taxonomy and Algorithmic Assessment of Wikipedia's Verifiability

Miriam Redi, Besnik Fetahu, Jonathan Morgan and Dario Taraborelli



WIKIMEDIA
FOUNDATION

Our Research on Content Integrity.



- Identify and address threats to content integrity
- Verify knowledge in the absence of secondary sources
- Characterize the integrity of multimedia knowledge
- Help contributors monitor violations of core content policies and assess information reliability
 - **Help editors find sentences missing citations**
 - Understand, structure and improve citations and their usage

Wikipedia:Verifiability

In **Wikipedia**, **verifiability** means other people using the encyclopedia can check that the information comes from a **reliable source**. All material in **Wikipedia mainspace**, including everything in articles, lists and captions, must be verifiable. All quotations, and any material whose verifiability has been challenged or is likely to be challenged, must include an **inline citation** that directly supports the material



Key to verifiability: presence of reliable sources

Reliable source: not
self-published research,
blogs, etc

Russian involvement

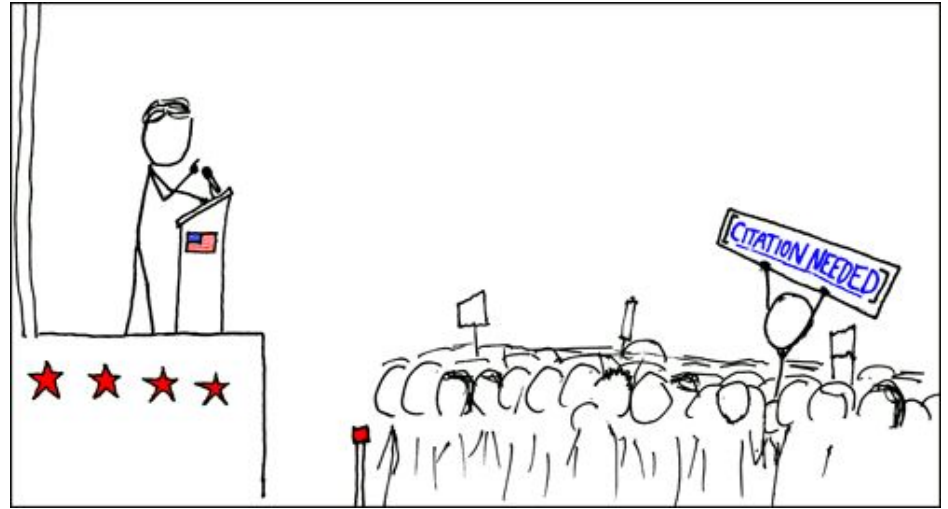
Vladimir Putin involvement

We assess Russian President Vladimir Putin ordered an influence campaign in 2016 aimed at the US presidential election. Russia's goals were to undermine public faith in the US democratic process, denigrate Secretary Clinton, and harm her electability and potential presidency. We further assess Putin and the Russian Government developed a clear preference for President-elect Trump. We have high confidence in these judgments.^[4]

This assessment^[4] was published in public, non-classified form in January 2017 by the Office of the Director of National Intelligence representing the work of the Federal Bureau of Investigation (FBI), the Central Intelligence Agency (CIA), and the National Security Agency (NSA). The FBI and CIA gave the assessment with high confidence and the NSA with moderate confidence.

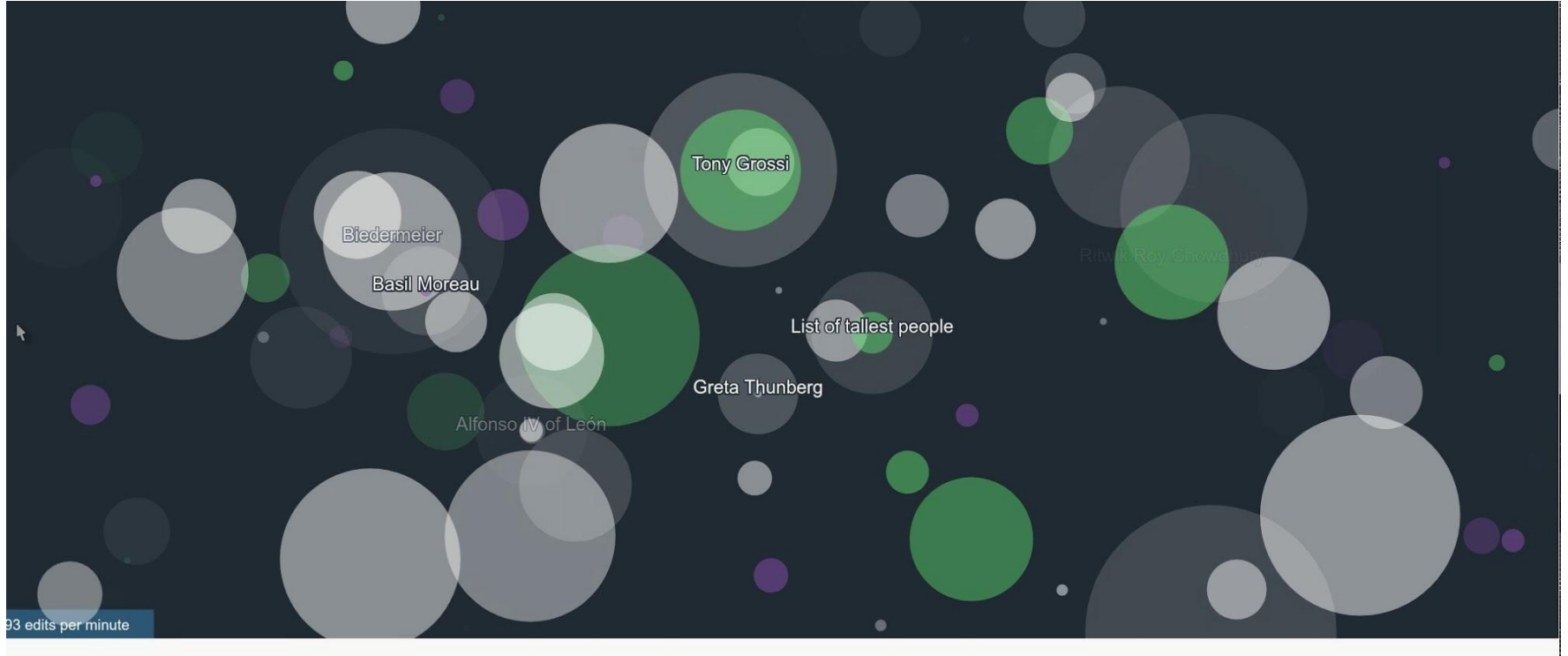
Citation needed

{{Citation needed}} template is manually added by editors to signal that the reader should read the content with care, and also that help is welcome to support the statement.



{{Citation needed|reason=Your explanation here|date=September 2019}}

Recent Changes



**Can we help editors identify *if* and
why Wikipedia statements need
citations?**

(using machine learning)

Wikipedia:Verifiability

In [Wikipedia](#), **verifiability** means other people using the encyclopedia can check that the information comes from a [reliable source](#). All material in [Wikipedia mainspace](#), including everything in articles, lists and captions, must be verifiable. All quotations, and any material whose verifiability **has been challenged or is likely to be challenged**, must include an [inline citation](#) that directly supports the material

What needs to be cited?

Wikipedia:Policies and guidelines

From Wikipedia, the free encyclopedia

Citation Reason Taxonomy

Detecting
Citation Need



Detecting
Citation
Reason

“if”

“why”

Citation Reason Taxonomy: Requirements

- A systematic definition of what needs citations and what doesn't
- Needs to be grounded in needs and rules of the Wiki Community
- Needs to be multilingual (beyond english-only)

Citation Reason Taxonomy: Problems

- Citation guidelines are not “exact”

Wikipedia: You **don't need to**
cite that the sky is blue

Wikipedia: You **do need to**
cite that the sky is blue

- Citation Guidelines might not be universal across languages

Citation Reason Taxonomy: Methods

- Qualitative analysis and understanding of citation guidelines across 3 languages – Italian, French and English
- Multilingual User Study with data annotation

Citation Reason Taxonomy:

Studying, analyzing and clustering citation policies from English, French and Italian Wikipedias

Category	Rule	Needs Cit?	References
Objective Cases	Direct Quotations	Y	<ol style="list-style-type: none"> https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Lead_section#Citations https://en.wikipedia.org/wiki/Wikipedia:Inline_citation#When_you_must_use_inline_citations https://en.wikipedia.org/wiki/Wikipedia:Verifiability#Responsibility_for_providing_citations https://en.wikipedia.org/wiki/Wikipedia:Citing_sources#Dealing_with_unsourced_material https://en.wikipedia.org/wiki/Wikipedia:Citing_sources#When_and_why_to_cite_sources https://en.wikipedia.org/wiki/Wikipedia:Likely_to_be_challenged https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view#Controversial_subjects https://en.wikipedia.org/wiki/Wikipedia:When_to_cite https://en.wikipedia.org/wiki/Template:Citation_needed https://en.wikipedia.org/wiki/Wikipedia:Citation_needed
	Statistics and Data	Y	
	Published Opinion	Y	
	Close Paraphrasing	Y	
Likely To Be Challenged	Similar statements in the past have been tagged as 'citation needed'	Y	
	Similar statements in the past have been removed or discussed for sourcing reasons	Y	
Controversial Facts (see [13] for a list of controversial issues)	Facts from Fringe and Conspiracy theories	Y	
	Facets of identity outlined at en:WP:CATEGRS	Y	
	Surprising Claims (not widely accepted)	Y	
	Tergiversation (Weasel Words) - complete list of words to watch in [15]	Y	
	Statements about a Living Person, which indicate either positive or negative action/opinion		
	Personal information about a living person, including details such as exact date of birth, relationship status, ethnicity, etc		
Technical Knowledge	Scientific fact	Y	<ol style="list-style-type: none"> https://en.wikipedia.org/wiki/Wikipedia:Verifiability#Exceptional_claims_require_exceptional_sources https://en.wikipedia.org/wiki/Wikipedia:Common_knowledge https://en.wikipedia.org/wiki/Wikipedia:List_of_controversial_issues https://en.wikipedia.org/wiki/Wikipedia:Citation_underkill https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Words_to_watch#Contentious_labels https://en.wikipedia.org/wiki/Wikipedia:What_the_Good_article_criteria_are_not#282.29_Factually_accurate_and_verifiable https://en.wikipedia.org/wiki/Wikipedia:Biographies_of_living_persons#Reliable_sources
	Medical claims (very dense citations needed)	Y	
	Statement on X being legal/illegal	Y	
	Historical claims (not common knowledge)	Y	
	Linguistic usage (not common knowledge)	Y	
Common Knowledge	Established historical fact (Cesar was a Roman)	N	
	Geographic information you can find on google maps	N	
	Observable Facts (Pisa Tower is Leaning)	N	
	National Associations (Italians Speak Italian)	N	
	Math truism	N	
	Elementary Knowledge on counts, time, etc	N	

Citation Reason Taxonomy:

Studying citation need with English, French and Italian editors

WikiLabel

Workset

3

Please read the highlighted statement and judge whether a citation is needed, following these [guidelines](#)... Please also provide a reason for your choice using the free text box.

Please use all the necessary contextual information (article title, position in article, surrounding text, ...) to help your decision.

write reason here [required]

Intrapersonal phenomena [edit source]

Attitudes [edit source]

Main article: *Attitude (psychology)*

In social psychology, attitudes are defined as learned, global evaluations of a person, object, place, or issue that influence thought and action.^{[1][page needed]} Put more simply, attitudes are basic expressions of approval or disapproval, favorability or unfavorability, or as them put it, likes and dislikes.^[1] Examples would include liking chocolate ice cream, or endorsing the values of a particular political party.

social psychologists have studied attitude formation, the structure of attitudes, attitude change, the function of attitudes, and the relationship between attitudes and behavior. Because people are influenced by the situation, general attitudes are not always good predictors of specific behavior. For example, for a variety of reasons, a person may enter the environment but not receive a view on a particular day.

In recent times, research on attitudes has examined the distinction between traditional, self-reported attitude measures and "implicit" or unconscious attitudes. For example, experiments^[citation needed] using the Implicit Association Test have found that people often demonstrate implicit bias against other races, even when their explicit responses reveal equal mindedness. One study found that explicit attitudes correlate with verbal behavior in interracial interactions, whereas implicit attitudes correlate with nonverbal behavior.^[12]

One hypothesis on how attitudes are formed, first advanced by Abraham Tesser in 1963, is that strong likes and dislikes are rooted in our genetic make-up. Tesser speculates that individuals are disposed to hold certain strong attitudes as a result of inborn physical, sensory, and cognitive skills, temperament, and personality traits. Whatever disposition nature elects to give us, our most treasured attitudes are often formed as a result of exposure to attitude objects; our history of rewards and punishments; the attitude that our parents, friends, and enemies express; the social and cultural context in which we live; and other types of experiences we have. Obviously, attitudes are formed through the basic process of learning. Numerous studies have shown that people can form strong positive and negative attitudes toward neutral objects that are in some way linked to emotionally charged stimuli.^{[13][185-186]}

Attitudes are also involved in several other areas of the discipline, such as conformity, interpersonal attraction, social perception, and prejudice.

Does the sentence need a citation?

Why?

Article with highlighted sentence

Wikilabels experiment translated in 3 languages

Citation Reason Taxonomy: Final Taxonomy

Reasons for adding a citation^[edit]

- The statement appears to be a direct **quotation** or close paraphrase of a source
- The statement contains **statistics** or data
- The statement contains surprising or potentially **controversial** claims - e.g. a **conspiracy** theory
- The statement contains claims about a person's subjective **opinion** or idea about something
- The statement contains claims about a person's **private life** - e.g. date of birth, relationship status.
- The statement contains technical or **scientific** claims
- The statement contains claims about general or **historical** facts that are not common knowledge

Reasons for not adding a citation^[edit]

- The statement only contains **common knowledge** - e.g. established historical or observable facts
- The statement is in the **lead section** and its content is referenced elsewhere in the article
- The statement is about a **plot** or character of a book/movie that is the main subject of the article
- The statement only contains claims that have been **referenced elsewhere** in the paragraph or article

Citation Reason Taxonomy

Detecting
Citation Need



Detecting
Citation
Reason

“if”

“why”

Citation Need Task:

does this statement need a citation? A binary classification task.

Data Collection

Data Modeling

Evaluation and Result Interpretation

Citation Need Task: Data Collection

English Wikipedia

But models are now ready for French and Italian too

Binary labels

POSITIVE examples:

10,000 statements with a citation

NEGATIVE examples:

10,000 statements without a citation

3 Article Sources to test generalizability:

FEATURED articles:

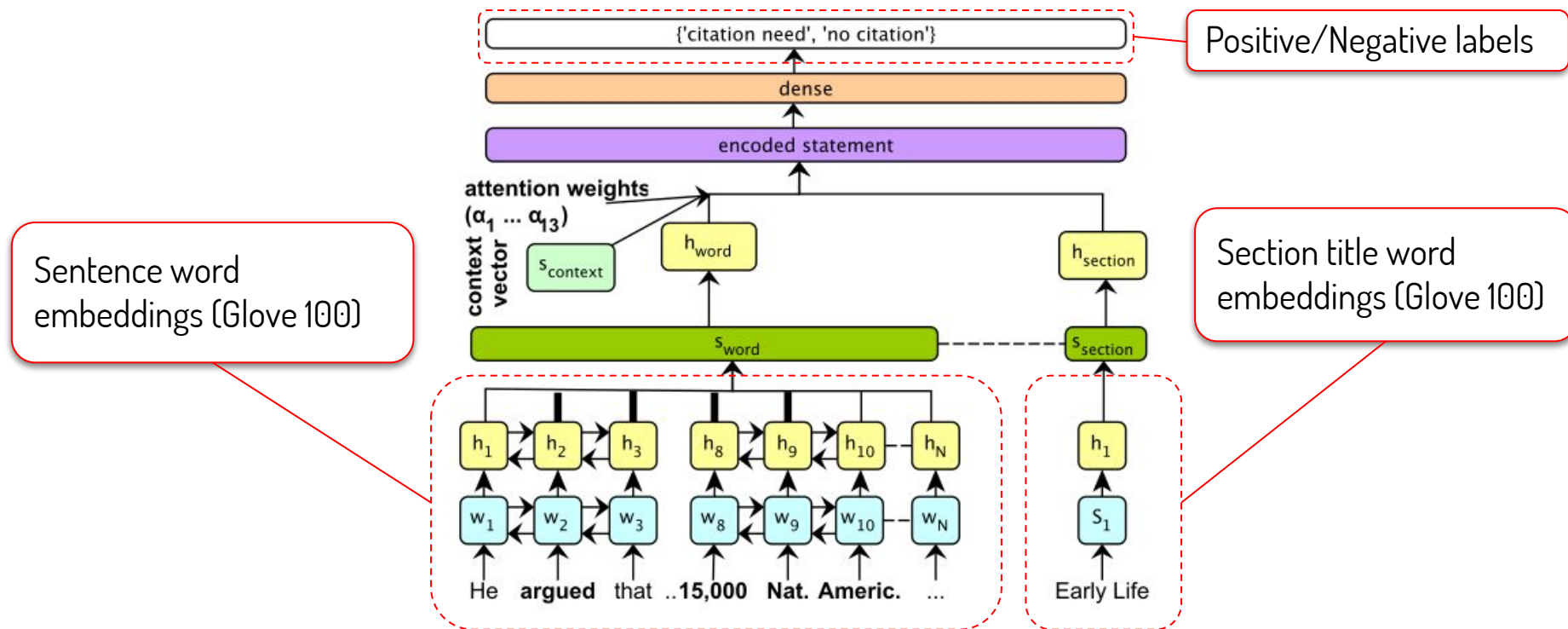
Best articles in
Wikipedia

LOW QUALITY articles.

Articles missing citations -
positives statements with
citation needed tag.

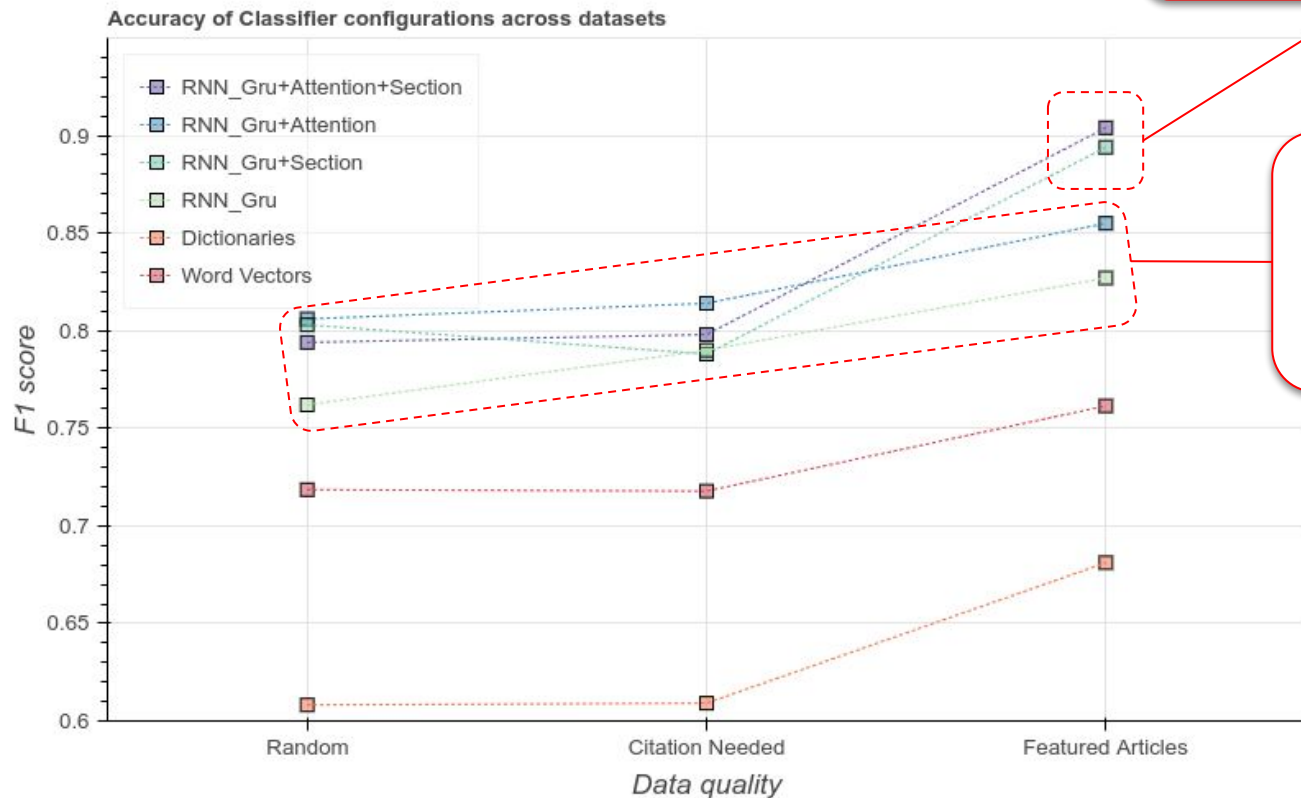
RANDOM articles. Articles
of varying quality and
topics randomly sampled
from Wikipedia

Citation Need Task: Data Modeling



Citation Need Task: Model Accuracy

Section information is important for Featured articles



Accuracy in general is substantially high across datasets (up to 90% for FA with Section information, 83% without Section info)

Citation Need Task: Analysis

Factual and assertive verbs are among the most highly related with citation need – section is very important for well-designed articles

FA		LQN		RND	
<i>Section</i>	-0.621	<i>underline</i>	0.054	<i>say</i>	0.084
<i>say</i>	0.107	<i>say</i>	0.0546	<i>underline</i>	0.0842
<i>underline</i>	0.107	<i>believe</i>	0.042	<i>Section</i>	-0.072
<i>realize</i>	0.068	<i>disagree</i>	0.040	<i>report</i>	0.062
<i>suggest</i>	0.068	<i>claim</i>	0.039	<i>tell</i>	0.062

Table 3: Point-Biserial Correlation Coefficient between citation need labels and individual feature values

Citation Need Task: Model's attention weights

Statistics

it is **estimated** that throughout **florida** the **storm** damaged 101 241 homes and destroyed approximately 63 000 others the vast majority in dade county with about 175 000 people rendered homeless

Scientific

in this setting the mathematics describing the gravitational field simplifies drastically and one can study **quantum** gravity **using** familiar methods from quantum field theory **eliminating** the need for string theory or other more radical approaches to quantum gravity in four dimensions

Life

writer paul dini **batman** : the **animated** series **detective** comics was first approached by dc **comics** around late 2007 about the prospect of creating a story for an original batman video game

Opinion

claimed the humor **kept** them **plowing** through the game regardless of the **issues** they encountered and were disappointed when the game steadily lost this approach in the later stages

History

in the 18th - century king charles iii of spain commissioned anton **raphael** mengs to **paint** the **triumph** of trajan on the ceiling of **the** banquet hall of the royal palace of madrid considered among the best works of this artist

Quotation

writer peter **hastings** said that **he** unintentionally **created** these catchphrases when he wrote the **episode** win big and then producer sherri stoner used them and had them put into later episodes

Controversial

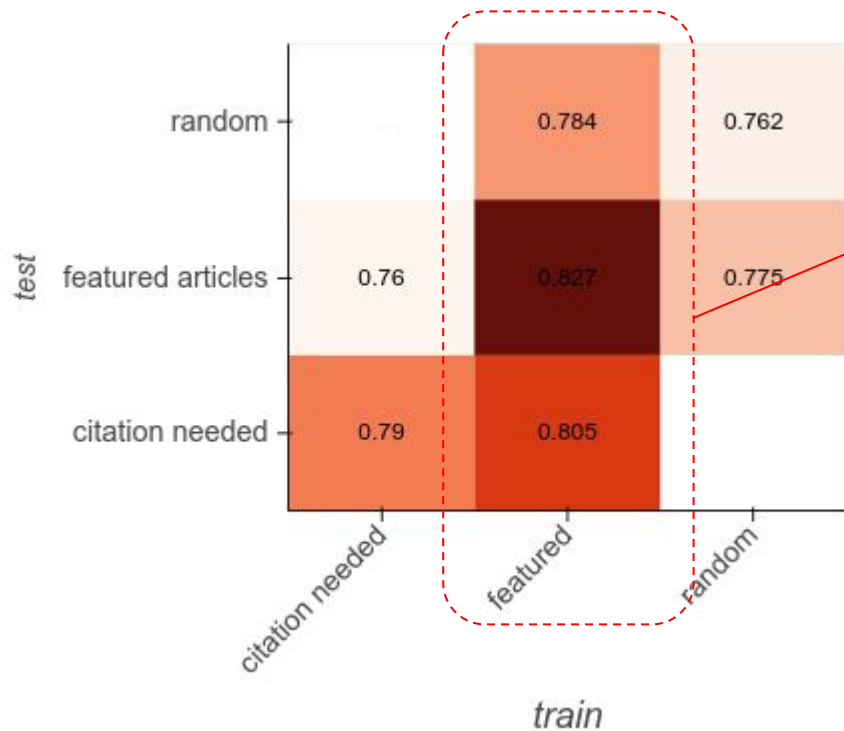
victoria withdrew from **public** life and **her** seclusion **eroded** some of albert work in attempting **to** re - model the monarchy as a national institution setting a moral if not political example

Other

this theorem is ultimately connected **with** the spectral **characterization** of **as** the eigenvalue associated with the heisenberg **uncertainty** principle and the fact that equality holds in the uncertainty principle only for the gaussian function

Again factual verbs + weasel words + domain-specific terms

Citation Need Task: Model's Generalizability



Model trained on featured articles generalizes across datasets

Citation Reason Taxonomy

Detecting
Citation Need



Detecting
Citation
Reason

“if”

“why”

Citation Reason Task:

Why does this statement need a citation? A multiclass classification task.

Taxonomy Building -- done!

Data Collection

Evaluation and Result Interpretation

Citation Reason Task: Data Collection

MTURK experiment: why does this statement need a citation?

Wikipedia Article: [Norwich City F.C.](#)

Sentence to be categorized:

Actor Hugh Jackman is also a fan of the club, having been taken to Carrow Road as a child by his English mother, though he turned down an opportunity to become an investor in the club in 2010

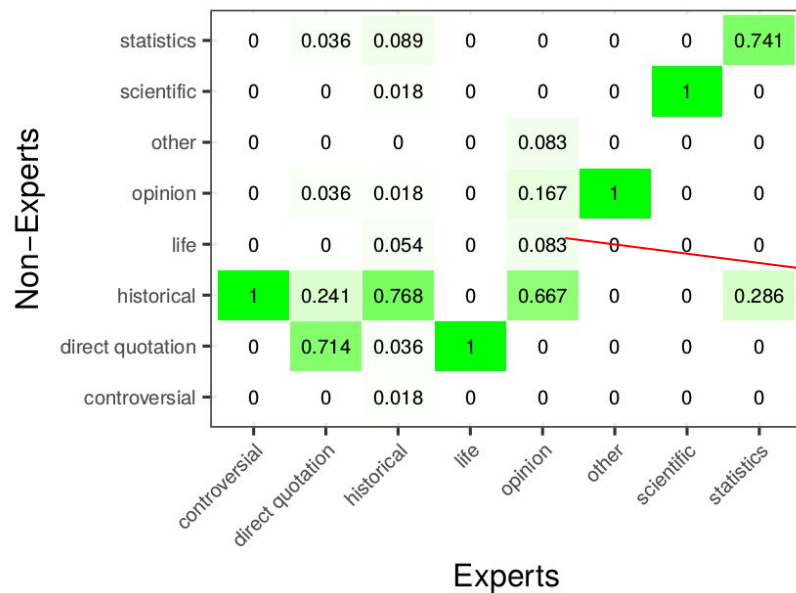
4000 sentences with citations from Featured articles

Choose a category:

Quotation

8 categories to choose between - mutually exclusive

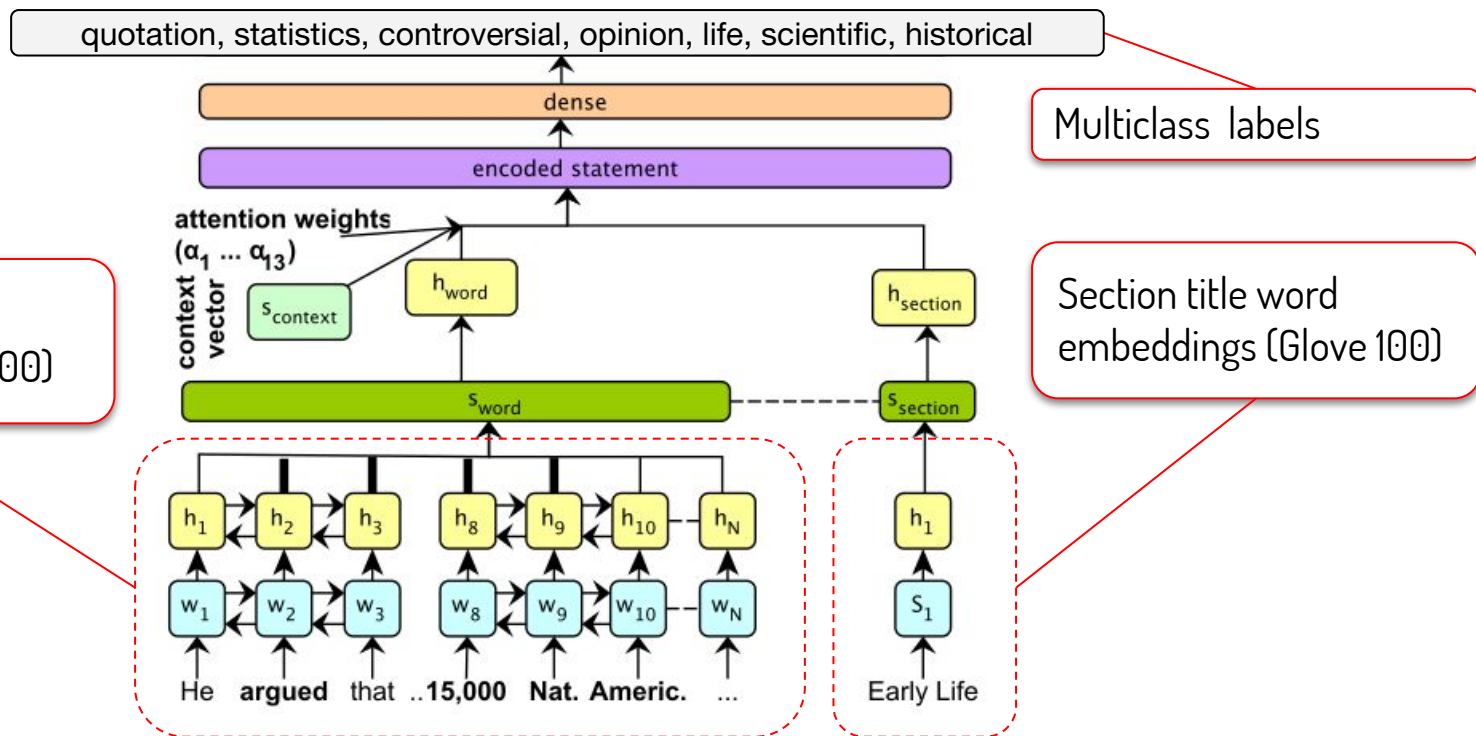
Citation Reason Task: Data Collection Results



Agreement is generally high, but we analyzed the cases where agreement is lower, and actually discover that reason classes are not mutually exclusive!

Figure 3: Confusion matrix indicating the agreement between Mechanical Turk workers ("non-experts") and Wikipedia editors ("experts"). The darker the square, the higher the percent agreement between the two groups

Citation Reason Task: Data Modeling, network finetuning

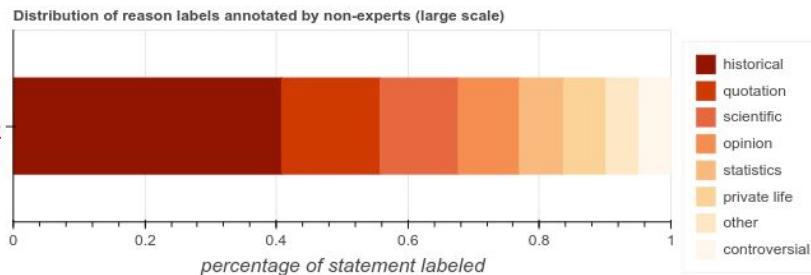


Citation Reason Task: Classification Results

	<i>pre-trained</i>			<i>no pre-training</i>		
	P	R	F1	P	R	F1
<i>direct quotation</i>	0.44	0.65	0.52	0.43	0.46	0.45
<i>statistics</i>	0.20	0.20	0.20	0.28	0.15	0.19
<i>controversial</i>	0.12	0.02	0.04	0.04	0.01	0.02
<i>opinion</i>	0.20	0.12	0.15	0.19	0.12	0.15
<i>life</i>	0.13	0.06	0.09	0.30	0.06	0.10
<i>scientific</i>	0.62	0.56	0.59	0.54	0.58	0.56
<i>historical</i>	0.56	0.67	0.61	0.54	0.74	0.62
<i>other</i>	0.13	0.05	0.07	0.14	0.08	0.10
avg.	0.30	0.29	0.28	0.31	0.28	0.27

Table 6: Citation reason prediction based on a pre-trained RNN_a^{+S} model on the FA dataset, and a RNN_a^{+S} which we train only on the citation reason dataset.

Accuracy is higher for classes where we have higher number of data examples



WHAT'S NEXT?

Citation Reason Taxonomy

Detecting
Citation Need

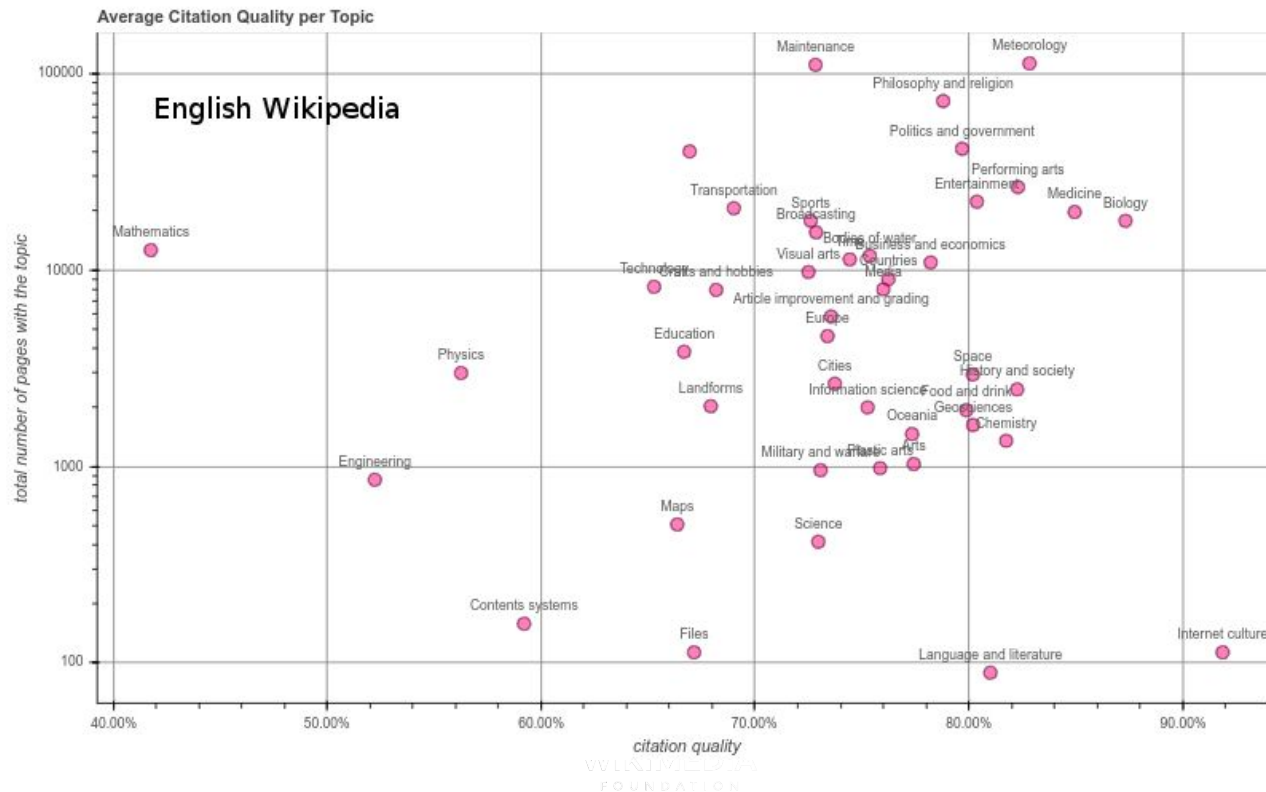


Detecting
Citation
Reason

“if”

“why”

Understanding citation quality at scale



Citation Needed Template Recommendation API

Research:Identification of Unsourced Statements/API design research



< [Research:Identification of Unsourced Statements](#)



This page documents a [research project in progress](#).

Information may be incomplete and change as the project progresses.
Please contact the project lead before formally citing or reusing results from this page.

Tracked in [Phabricator](#)
[Task T228816](#)

Created

21:59, 31 July 2019 (UTC)

Contact

[Jonathan Morgan](#)
Wikimedia Foundation

Collaborators

[Miriam Redi](#)
Wikimedia Foundation

Duration: 2019-August
— ??

≡ [Contents](#) ▾

^ Research goals



[Research:Projects](#)

This research project is focused on identifying design requirements for a service to make the classifications of the Unsourced Statement models available to volunteer tool developers within the Wikimedia Movement. We hope that these tool developers, in turn, develop tools (e.g. bots, gadgets, web applications) that editors can use to identify and correct unsourced statements.

Thank you 🧐

miriam@wikimedia.org

@mad_astronaut

Project page: https://meta.wikimedia.org/wiki/Research:Identification_of_Unsourced_Statements

Code: <https://github.com/mirrrys/citation-needed-paper/>

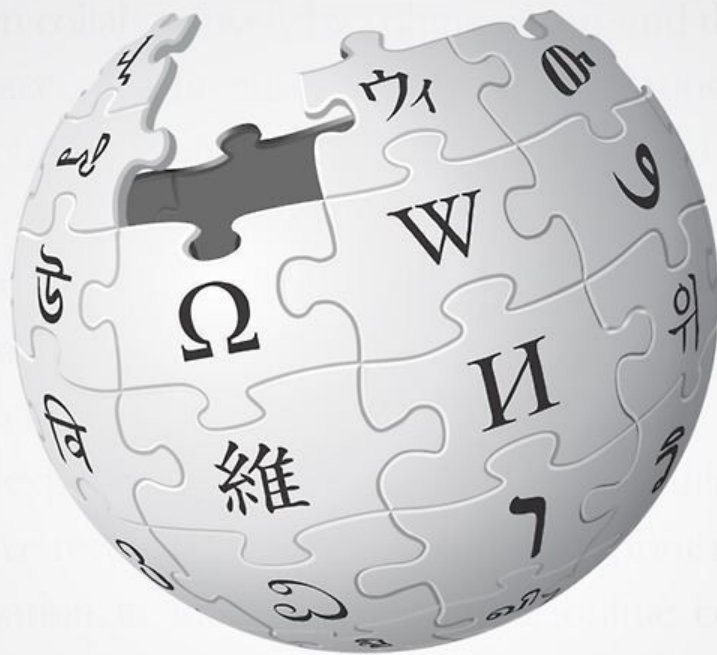
Data: https://figshare.com/articles/Citation_Reason_Dataset/7756226

Paper: <https://arxiv.org/pdf/1902.11116.pdf>



WIKIMEDIA
FOUNDATION

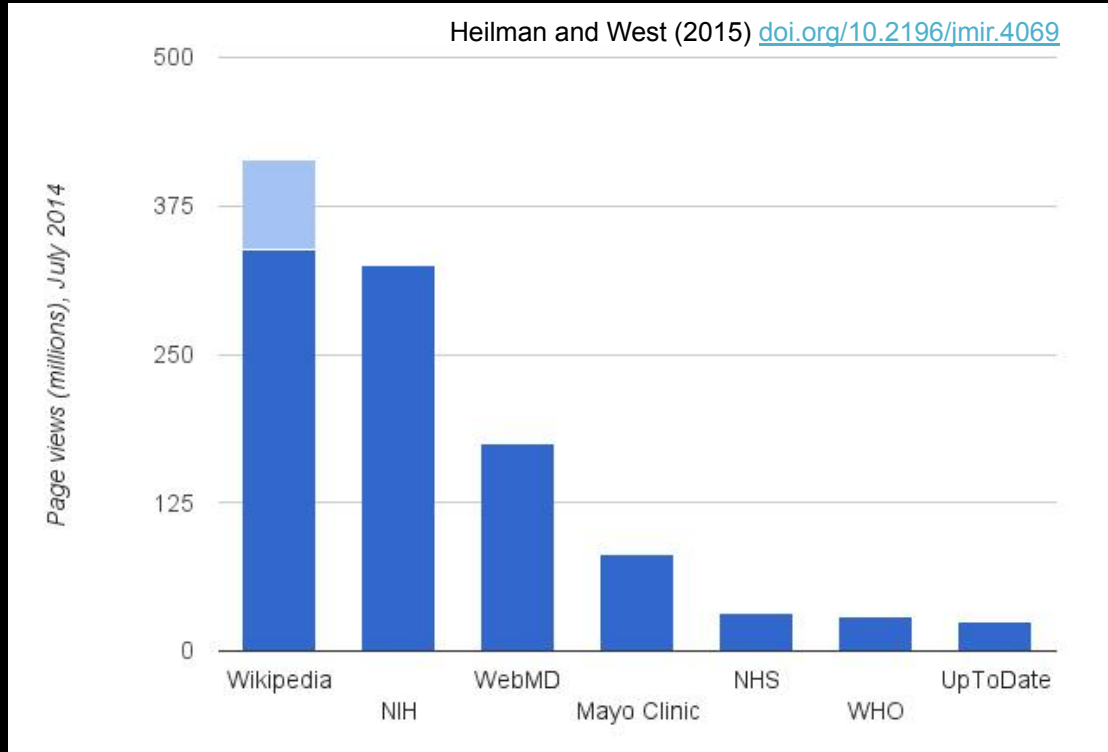
A major entry point into the scholarly literature ...



WIKIPEDIA

You're welcome, college students.

The world's most visited online medical resource



MENU ▾

nature
International journal of science

News | Published: 14 Dec

Internet er

Jim Giles

Nature 438, 900–901 (15 D

Jimmy Wales' W
accuracy of its s
see details of ho
supplementary ;
reported in this

Do
He
Wil

Fifty per
are edit
informa



Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store
Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page
Tools
What links here
Related changes
Upload file
Special pages
Permanent link

Article Talk

Read

View source

View history

Search Wikipedia

🔍

Science

From Wikipedia, the free encyclopedia

This article is about the general term. For other uses, see Science (disambiguation).

Science (from Latin *scientia*, meaning "knowledge"^{[1][2][58]} is a systematic enterprise that builds and organizes knowledge in the form of testable explanations and predictions about the universe.^[4]

Contemporary science is typically subdivided into the natural sciences, which study the material universe; the social sciences, which study people and societies; and the formal sciences, which study logic and mathematics. The formal sciences are often excluded as they do not depend on empirical observations.^[3] Disciplines which use science, like engineering and medicine, may also be considered to be applied sciences.^[4]

From classical antiquity through the 19th century, science as a type of knowledge was more closely linked to philosophy than it is now, and in the Western world the term "natural philosophy" once encompassed fields of study that are today associated with science, such as astronomy, medicine, and physics.^[59] However, during the Islamic Golden Age foundations for the scientific method were laid by Ibn al-Haytham in his Book of Optics.^{[16][20][92][93]} While the classification of the material world by the ancient Indians and Greeks into air, earth, fire and water was more philosophical, medieval Middle Easterners used practical and experimental observation to classify materials.^[11]

In the 17th and 18th centuries, scientists increasingly sought to formulate knowledge in terms of physical laws. Over the course of the 19th century, the word "science" became increasingly associated with the scientific method itself as a disciplined way to study the natural world. It was during this time that scientific disciplines such as biology, chemistry, and physics reached their modern shapes. That same time period also included the origin of the terms "scientist" and "scientific community", the founding of scientific institutions, and the increasing significance of their interactions with society and other aspects of

Not logged in | Talk | Contributions | Create account | Log in

We have a new blockchain news roundup newsletter, it's delightful

Forget what your school says, MIT research proves Wikipedia is a source for science

by TRISTAN GREENE — 1 year ago in INSIGHTS

Part of a series on
Science

Formal [show]

Physical [show]

Life [show]

Social [show]

Applied [show]

Interdisciplinary [show]

Philosophy · History [show]

Outline · Portal · Category

V · T · E