The discrepancy in Credibility Ratings over Manual Fact-Checking of Misinformation

Saeedeh Shekarpour, Rushabh Phadkule, Aditya Narvekar, Nileena Thomas
University of Dayton, USA
{sshekarpour1,phadkuler1, narvekara1, thomasn10}@udayton.edu

Abstract

Manual Fact-checking provides a precious body of content which is useful for building AI applications over various tasks such as source-checking, authenticity-checking, stance detection, and style-checking. However, the current state of the reports published by fact-checkers is not friendly for machine processing since there is no unified agreement upon organizing, structuring, and annotating reports. In this paper, we focus on the credibility rating systems employed by various fact-checking organizations. These ratings do not follow any systematic principle with respect to granularity, scale or even computation. Such discrepancies bring hurdles for integrating and consuming content across fact-checking organizations. Herein, we contribute to identify and analyze the credibility labeling systems utilized by around 11 active fact-checking organizations. Then, we introduce ontological concepts that categorize, integrate and reorder the labels in order to reach a best practice scale over credibility ratings.

1. Introduction

The increasing rate of information pollution (DesertNews, 2019; IDEAS, 2019) on the Web requires novel solutions to tackle. According to the recent Pew research in 2019 (PewResearch, 2019), Americans rated misinformation more critical than identifying terrorism, illegal immigration, racism, sexism, even climate change. Such facts indicate several major deficiencies in the area of computation, information, and Web science. Although the research and technology communities try to combat the misinformation challenge using fully automatic approaches (based on Artificial Intelligence (AI) models), misinformation is a complex and multifaceted problem (Ananny, 2019). Since decades volunteer and professional fact-checkers have been debunking the news, yet there has not been any indexing, integrating, interlinking over the content generated by them. In 2015, the International Fact-Checking Network (IFCN) was founded as a unit of the Poynter Institute[1]. It gathers fact-checkers worldwide to introduce best practices for fact-checking initiatives. The trained journalists or fact-checkers spend hours or even days to approve or disapprove a given piece of information (e.g., a news article). They go through rigorous steps to verify various parts of a given claim. Their investigations lead to a published report containing a precious body of content. This report is also often concluded (assigned) a credibility rating (e.g., `Mostly True`, `True`, `False`, `Mostly False`, `Hoax`, `Satire`, `Unproven`, `Outdated`, `Uncertain`, `0 Crows`, `1 crow`) indicating the overall assessment on the claim.

Manual Fact-checking provides a precious body of content which is useful for building up AI applications over various tasks such as source-checking, authenticity-checking, stance detection, and style-checking. However, the current state of the reports published by fact-checkers is not friendly for machine processing since there is no unified agreement upon organizing, structuring, and annotating reports. Although recently `schema.org`[2] extended a standard schema called `ClaimReview`[3] for publishing the report of fact-checkers in a structured format. However, so far, only a few fact-checkers have adopted it[4]. In this context, we focus only on the credibility ratings utilized in various fact-checking organizations. We represent the discrepancy between them. Such discrepancies bring hurdles for integrating and consuming content across fact-checking organizations. Herein, we contribute to identify and analyze the credibility labeling systems utilized by around 11 active fact-checking organizations. Then, we introduce ontological concepts that categorize, integrate and re-order the labels in order to reach a best practice scale over credibility rating. This paper is organized as follows: Section 2 discusses the concept of discrepancy in the domain of debunking information (i.e., fact-checking). Then, in Section 3, we represent the 11 fact-checking websites that we analyzed along with the demonstration of their credibility rat-

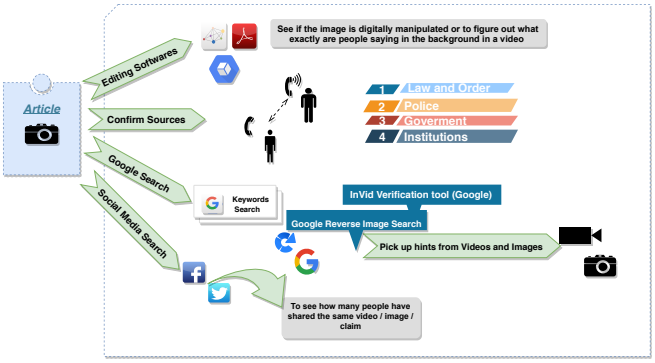

Figure 1: The overall Methodology used to rate the claims by `Times of India`.

[1] `https://www.poynter.org/`

[2] `http://schema.org/`

[3] `https://schema.org/ClaimReview`

[4] `https://reporterslab.org/a-better-claimreview-to-grow-a-global-fact-check-database/`

Website	Owned By/Founder	Category	URL
SNOPE	Snopes Media Group	Independent Publication	https://www.snopes.com/
Politifact	Poynter Institute	Non-Partisan Website	https://www.politifact.com/
Truth or Fiction	Whats True Incorporated	Non-Partisan Website	https://www.truthorfiction.com/
Washington Post	Jeff Bezos	Multi-Partisan,Media Website	https://www.washingtonpost.com/
Vishvas News	MMI Online Limited	Non-Partisan,Media Website	https://www.vishvasnews.com/
Teyit	Teyit Media Research Association	Non-Partisan Website	https://teyit.org/
Africa Check	AFP	Non-Partisan,Non-Profit Website	https://africacheck.org/
Peoples Pundit Daily	PPD Ventures LLC	Non-Partisan Website	https://www.peoplespunditdaily.com/
India Today	Vidya Vilas Purie, Madhu Trehan, and Aroon Purie	Media, Journalism Website	https://www.indiatoday.in/
Check Your Fact	The Daily Caller, Inc	Non-partisan,Independent Website	https://checkyourfact.com/
Polygraph	Voice of America and Radio Free Europe/Radio Liberty	Non-partisan,International Broadcasters	https://www.polygraph.info/
Emergent	Craig Silverman	Research Project	http://www.emergent.info/

Table 1: **Fact-checking Websites and their categorization:**

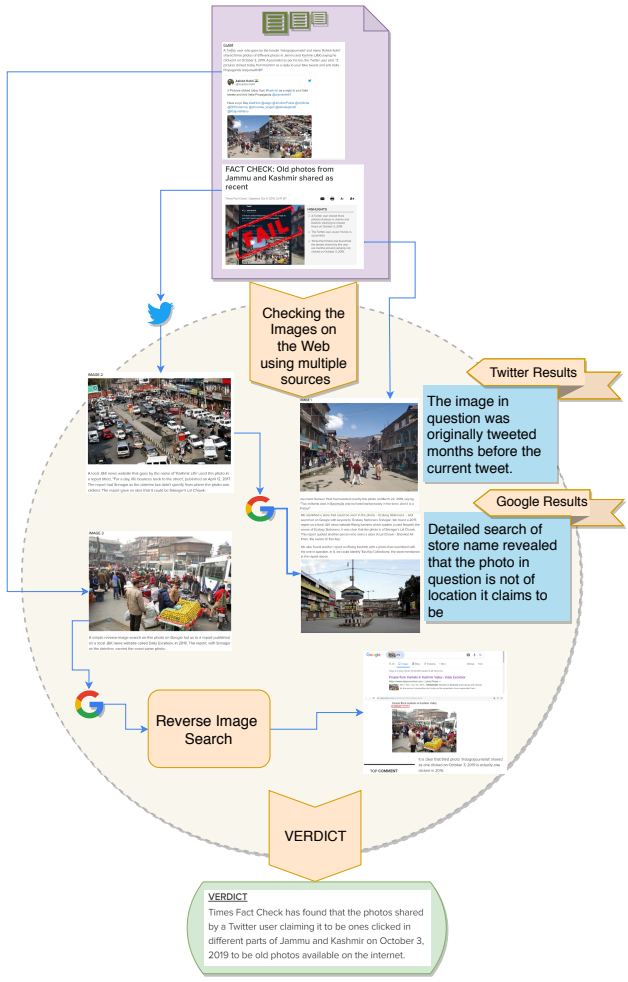

Figure 2: An example of an applied methodology for debunking information in Times of India.

ings. Next, we provide a categorization over the credibility ratings in general in Section 4. We close with conclusive remarks and our future plan.

2. Discrepancy

The manual debunking of information is inherently subjective. In order to commit to transparency, IFCN has set up a code of principles[5]. However, there are still an inherent discrepancies in the employed methodologies, published reports, or conclusive ratings. The work presented in (Mensio and Alani, 2019) recognizes this and tries to resolve that using an aggregation over a Web of Trust Network (Golbeck et al., 2003; Shekarpour and Katebi, 2010). In the following, we shortly discuss the discrepancy in the methodology and credibility ratings.

Discrepancy in Methodology. We carefully investigated and observed the 11 Fact-Checking websites listed in Table 1. We concluded that all of them rely on discrepant (i) methodological tools to review and judge the news, (ii) credibility scale and labels to express the conclusive opinion, (iii) structure of the report to detail their investigation. For example, Figure 1 and 2 demonstrate that Times of India employs different tools to approve or disapprove a given claim. They primarily check the sources of the claim by contacting the authorities such as the judicial system, police, involved governmental offices, or other involved organizations or authorities. Next, they search Google with the relevant keywords to find out related newspaper reports, published papers, or photographs. They also do search Images in the claim using Google reverse image search. This step helps them judge the misleading context or manipulations that might have happened to the original claim or images[6]. Checking social media (e.g., Twitter, Facebook) also leads to the indicators over the spread, acceptance, alteration, or manipulation of the claim by the public. Anyway, there is no standard or reference methodology for conducting the fact-checking procedure. Every organization or individual fact-checker might prioritize some strategies.

Discrepancy in Credibility Ratings. The process of assessing credibility often results in a conclusive label - (Mensio and Alani, 2019) names that as credibility value-, which boldly varies amongst various fact-checking websites. For example, SNOPE employs around 14 labels

[5] https://ifcncodeofprinciples.poynter.org/

[6] InVid Verification tool (Google extension) is an open-source video/image verification tool that is designed for journalists to obtain *"contextual information"* about Facebook/YouTube/Instagram/Twitter/Dailymotion videos. It also performs reverse-image search on Google, Baidu or Yandex search engines.

for concluding the credibility assessment, including labels such as `True, Mostly True`. While `India Today` relies on only four labels (i.e., `0 crow, 1 crow, 2 crows, 3 crows`, which are respectively equivalent to `True, Mixture, Mostly False, False` of `SNOPE`). Differently, `Truth or Fiction` has a binary labeling mechanism, and `Check Your Fact` website has a ternary labeling. Apart from the discrepancy in the number of labels, also there is a discrepancy in the identity of the labels. For example, among the 14 labels of `SNOPE` not all of them represent a degree of credibility. There are labels that indicate the status of the claim. For example, `Unproven` means, based on the current pieces of evidence, it can not be proved or disproved. Table 2 lists the labeling mechanism of our analyzed fact-checking websites along with a short description of them.

3. Fact-checking Organizations

We identified 11 fact-checking websites,among which are 10 listed in Table 1.(Polygraph.info does not mention characterization/definition of labels delegated to a claim, so we excluded it in the table) Furthermore, we analyzed the labeling systems employed by them to rate the credibility as the ultimate conclusion of their investigation. The set of all these labeling systems is represented in Table 2. In the following, we present a short introduction to each website.

Snopes: Snopes, launched in 1994, is an online fact-checking organization that focuses on debunking or validating urban legends and stories related to American Pop Culture. It is funded from the revenue generated through the advertisements on its website and donations from its users.

Politifact: Politifact is a non-profit project operated by the Poynter Institute for Media Studies in St. Petersburg, Florida. It is focused on operating on statements made by politicians and rating them for accuracy. It is funded by Tampa Bay Times newspaper and the revenue generated from content partnerships, online advertising, and grants. It was launched in 2007.

TruthOrFiction: TruthOrFiction.org is a website where users can procure information about rumors, fake news, misinformation, hoaxes, or stories circulated through email and social media. It is funded through programmatic digital advertisement sales and was launched in 1999.

Washington Post: Washington Post is a leading American daily newspaper published in Washington, D.C., USA. Nash Holdings, a holding company, established by Jeff Bezos, currently owns the Washington Posts. It has distinguished itself through political reporting on the functioning of the White House, Congress, and other departments of the U.S. Government. It was launched in 1877.

Vishwas News: Vishwas News is India's first dedicated Hindi fact-check and news verification portal, which is certified by the International Fact-Checking Network. It focuses on issues related to politics, sports, health, law, and justice, education, and science.

Teyit: Teyit is an independent fact-checking organization based in Turkey. Their main objective is to prevent the spreading of online false news. It is financially supported by the European Union and embassy funds granted to Non-Government Organisations and revenue generated from content partnerships and personal contributions. It was launched in 2016.

Africa Check: Africa Check is a non-profit organization founded by the media development arm of the international news agency, AFP. It focuses on topics related to crime and race in South Africa and fake healthcare information in countries around Africa. It was launched in 2012.

Peoples Pundit Daily: It delivers reader-funded data journalism on topics related to politics, elections, business, and economy. It is owned by PPD Ventures LLC and is funded through reader subscriptions, donations, and sponsored content. It was launched in 2013.

India Today: It delivers news related to politics, entertainment, economics, sports, and current affairs in India. It is owned by Living Media India Limited and launched in 1975.

Check Your Fact: It is a fact-checking news site that independently checks statements made by prominent people as well as news reported by other news outlets. It also checks for rumors which circulate on the internet. It is owned by The Daily Caller, Inc and is funded by the company's budget as well as revenue generated through online advertisements.

Polygraph.info: It is a fact-checking website that verifies the credibility of information circulated on the Internet. It is launched by Voice of America and Radio Free Europe/Radio Liberty. The organization is funded by the United States Government.

Emergent: Emergent is a real-time rumor tracker. It's part of a research project with the Tow Center for Digital Journalism at Columbia University that focuses on how unverified information and rumor are reported in the media. It aims to develop best practices for debunking misinformation.

4. Categorizing Credibility Labels

In this section, we introduce three ontological concepts that integrate and abstract all the labeling systems we presented in the previous section. This step is an initial step in a more extended agenda toward building a standard ontology for assessing credibility, assigning credibility rating. Our major aim is to overcome the discrepancy over the content generated by fact-checkers and represent that content in a more unified and homogeneous way. As we mentioned earlier, there is a bold discrepancy in the labeling systems employed to represent the credibility ratings. For example, `India today` considers four labels to rate the credibility of claims (i.e., `0 crow, 1 crow, 2 crows, 3 crows`). These sort of labels specify the degree of veracity or the falseness of a claim. However, there are labels that do not specify the degree of veracity. In fact, they categorize the type of a claim. For example the label `Hoax` or `Satire` from `SNOPE`. The third kind of label indicates the status of the credibility assessment. For example, `Checked`, `In-Checking` or `Proven` shows

Website	Label	Description
SNOPE	**True**	The primary elements of a claim are demonstrably true.
	Mostly True	The primary elements of a claim are demonstrably true, but some may be inaccurate
	Mixture	A claim has significant elements of both truth and falsity to it
	Mostly False	The primary elements of a claim are demonstrably false
	False	The primary elements of a claim are demonstrably false
	Outdated	Applies to items for which subsequent events have rendered their original truth rating irrelevant
	Miscaptioned	It is used with photographs and videos that are real but are nonetheless false
	Correct Attribute	Indicates that quoted material has been correctly attributed to the person who spoke or wrote it
	Misattributed	Indicates that quoted material has been incorrectly attributed to a person who didn't speak or write it
	Scam	This rating is not a truth rating but rather indicates pages that describe the details of verified scams
	Legend	Is most commonly associated with items that describe events so general or lacking in detail and are therefore essentially unprovable
	Labelled Satire	A claim is derived from content described by its creator and wider audience as satire
	Lost Legend	The Repository of Lost Legends (TRoLL for short) for those of who don't let the truth get in the way of a good story.
	Unproven	This indicates that insufficient evidence exists to establish the given claim as true, but cannot be definitively proved false
Politifact	**True**	The statement is accurate and there's nothing significant missing
	Mostly True	The statement is accurate but needs clarification or additional information
	Half True	The statement is partially accurate but leaves out important details or takes things out of context
	Mostly False	The statement contains an element of truth but ignores critical facts that would give a different impression
	False	The statement is not accurate
	Pants on Fire	The statement is not accurate and makes a ridiculous claim
	No Flip	No significant change in position
	Half Flip	A partial change in position
	Full Flip	A complete change in position
Truth or Fiction	**True**	Claims under observation are True Claims
	Not True	Claims under observation are False Claims
Washington Post	**One Pinocchio**	Some shading of the facts. Selective telling of the truth
	Two Pinocchio	Significant omissions and/or exaggerations. Some factual error may be involved
	Three Pinocchio	Significant factual error and/or obvious contradictions. This gets into the realm of mostly false
	Four Pinocchio	Totally False
	An Upside-Down Pinocchio	A statement that represents a clear but unacknowledged flip-flop from a previously-held position
	Bottomless Pinocchio	Four Pinocchios from The Fact Checker, and they must have been repeated at least 20 times
	The Geppetto Checkmark	Statements and claims that contain the truth, the whole truth, and nothing but the truth
	Verdict Pending	There are occasions when it is impossible to render a snap judgment because the issue is very complex or there are good arguments on both sides
Vishvas News	**True**	The news is true
	Fake News	The news is fake
	Misleading	The news misinforms
Teyit	**True**	It indicates that the data examined is true
	False	The data under review are false
	Mixed	It points out that the multiple propositions examined contain both true and false
	Uncertain	Data on the alleged claim were obtained; however, he states that these data are not sufficient to reach a conclusion
Africa Check	**Correct**	The claim is accurate, according to the best evidence publicly available at this time, and leaves out nothing significant
	Mostly Correct	The claim contains elements of truth but is either not entirely accurate,or needs clarification
	Unproven	Evidence publicly available at this time neither proves nor disproves the statement
	Misleading	Elements of the claim are accurate but presented in a way that it is misleading
	Exaggerated	The claim exaggerates the facts
	Understated	The claim understates the facts
	Incorrect	The claim is inaccurate according to the best evidence publicly available at this time
	Checked	Multiple claims have been fact-checked
Peoples Pundit Daily	**True**	Indicates the primary claim reviewed is verifiably true
	Mostly True	Indicates the primary claim reviewed is verifiably true, but peripheral or supporting relevant details could not be confirmed
	Half True	Indicates the primary claim reviewed has elements that are both true and false, so much so that it is only half true
	Mostly False	Indicates the primary claim reviewed is verifiably false, but peripheral or supporting relevant details can also be confirmed
	False	Indicates the primary claim reviewed is verifiably false
India Today	**0 Crow**	True
	1 Crow	Half True
	2 Crows	Mostly False
	3 Crows	Absolute False
Check Your Fact	**True**	The primary aspects of the claim are true and can be backed up with evidence
	False	The primary aspects of the claim are false and lack supporting evidence
	Unsubstantiated	There's not enough evidence to establish a claim as true or false
Emergent	**True**	The facet of the claim are true
	False	The facet of the claim are true
	Unverified	There's not enough evidence to establish a claim as true or false
	Controversial	The claims in question are arguably disputable

Table 2: **The Labeling Systems of 10 Fact-checking Websites**

whether the given claim has been assessed or it is still under assessment or if it is has been proven. Thus, we introduce three ontological class as follows:

1. `Credibility Type`: It represents a class of labels which show the type of credibility. This class is demonstrated in Figure 3. This class contains labels such as `Misleading, Hoax, Credible, Malinformation, Fact`. Figure 3 showcases a few of our mappings from the labeling systems to our class hierarchy.

2. `Credibility Status`: This class represents all the labels indicating the progress status of credibility assessment. Figure 4 shows this class and its associated sub-classes.

3. `Credibility Degree`: This class covers a discrete and continuous series of classes for representing the degree (value) that is assigned to an assessment. Figure 5 shows this hierarchy.

Please note that we represented only a few mappings from the credibility labeling to the proposed classes in Figure 3,4, and 5. We will have full representation within a mature ontology in upcoming future work. These categorizations can help us to organize the distinct labels of the various fact-checking organizations. Having a cohesive and standardized ontology for fact-checking will reduce ambiguity, discrepancy, and also pave the way for automatic integration of contents across various fact-checking websites.

5. Conclusion and Future Plan

In this paper, we presented our insights over the analysis of fact-checking websites. We recognized the discrepancy between labeling systems for rating credibility. To tackle that, we introduced three ontological concepts that differentiate and structure various labels. Our ultimate aim is to integrate the content from all fact-checking websites and represent them in a unified and standard way. Such representation will be beneficial for providing precious data for automating tools for fact-checking. This paper is an initial step towards a more extended agenda for developing a reference ontology for credibility assessment of fact-checkers.

6. Bibliographical References

Ananny, M. (2019). The partnership press: Lessons for platform-publisher collaborations as facebook and news outlets team to fight misinformation. `https://www.cjr.org/tow_center_reports/partnership-press-facebook-news-outlets-team-fight-misinformation.php`. Accessed: 2018-04-04.

DesertNews. (2019). Americans rate fake news as a bigger problem than terrorism. can twitter solve the problem? `https://www.deseretnews.com/article/900074903/fake-news-terrorism-twitter-social-media-president-donald-trump-kamala-harris-nancy-pelosi-pete-buttigieg-elizabeth-warren.html`. Accessed: 2019-09-11.

Golbeck, J., Parsia, B., and Hendler, J. (2003). Trust networks on the semantic web. In *International workshop on cooperative information agents*, pages 238–249. Springer.

IDEAS, D. A. G. (2019). Some real news about fake news. `https://www.theatlantic.com/ideas/archive/2019/06/fake-news-republicans-democrats/591211/`. Accessed: 2019-09-07.

Mensio, M. and Alani, H. (2019). Misinfome: Who's interacting with misinformation?

PewResearch. (2019). Many americans say made-up news is a critical problem that needs to be fixed. `https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/`. Accessed: 2019-09.

Shekarpour, S. and Katebi, S. (2010). Modeling and evaluation of trust with an extension in semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(1):26–36.

[7]

[7]

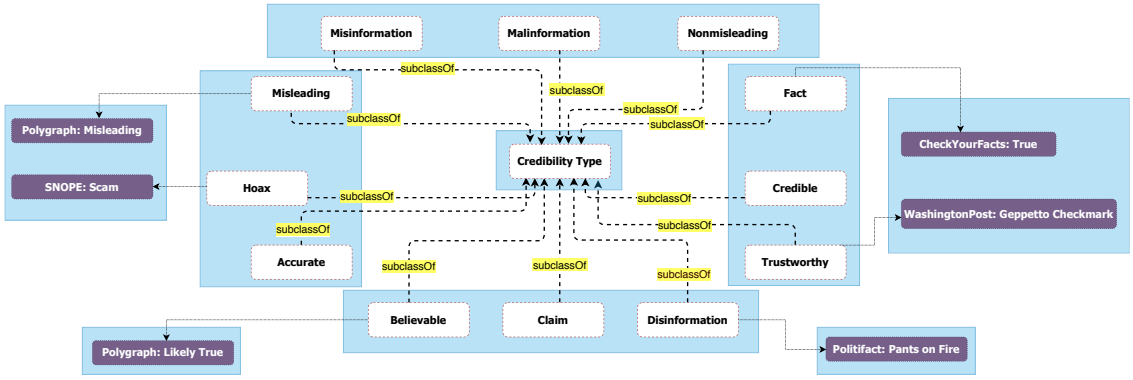

Figure 3: The class of `Credibility Type` and its sub-classes

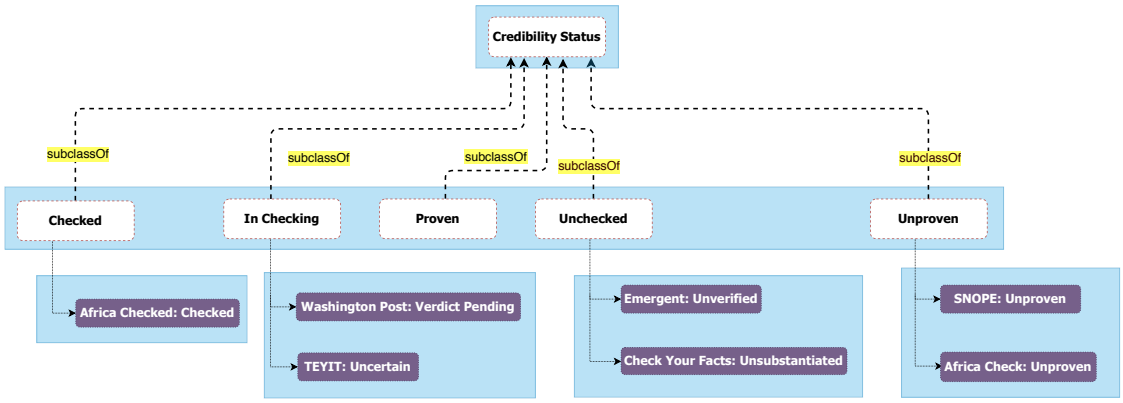

Figure 4: The class of `Credibility Status`.

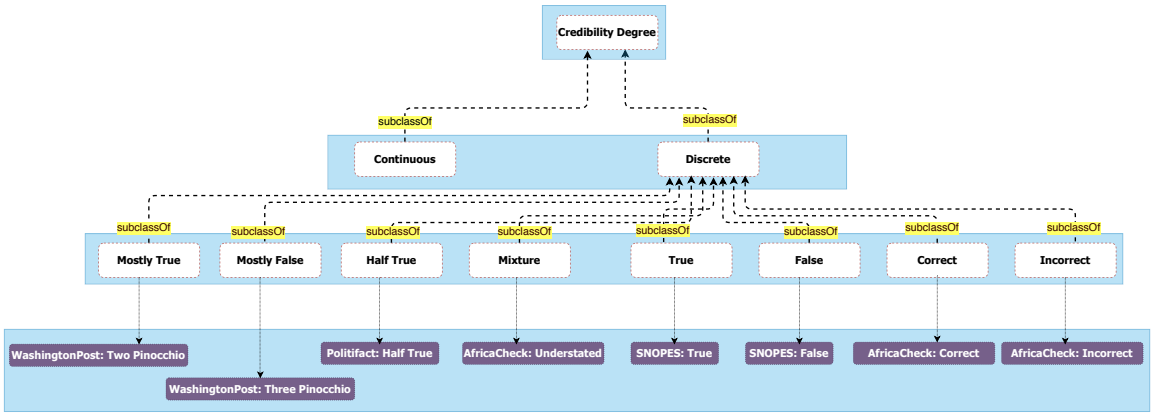

Figure 5: The class of `Credibility Degree`.