A Content Management Perspective on Fact-Checking

Sylvie Cazalens, Philippe Lamarre, Julien Leblay, Ioana Manolescu, Xavier Tannier

A Content Management Perspective on Fact-Checking

Sylvie Cazalens
Université de Lyon, CNRS
INSA-Lyon, LIRIS, UMR5205
Villeurbanne, F-69621, France
sylvie.cazalens@insa-lyon.fr

Philippe Lamarre
Université de Lyon, CNRS
INSA-Lyon, LIRIS, UMR5205
Villeurbanne, F-69621, France
philippe.lamarre@insa-lyon.fr

Julien Leblay
AIRC, AIST
Tokyo, Japan
julien.leblay@aist.go.jp

Ioana Manolescu
Inria, UMR7161
CNRS, Ecole Polytechnique
Palaiseau, 91120, France
ioana.manolescu@inria.fr

Xavier Tannier
Sorbonne Université, Inserm, LIMICS
Paris, France
xavier.tannier@sorbonne-universite.fr

ABSTRACT

Fact checking has captured the attention of the media and the public alike; it has also recently received strong attention from the computer science community, in particular from data and knowledge management, natural language processing and information retrieval; we denote these together under the term "content management". In this paper, we identify the fact checking tasks which can be performed with the help of content management technologies, and survey the recent research works in this area, before laying out some perspectives for the future. We hope our work will provide interested researchers, journalists and fact checkers with an entry point in the existing literature as well as help develop a roadmap for future research and development work.

1 INTRODUCTION

Journalism is about reporting the state of the world and events that affect society. Journalists' ethics aim to accuracy, objectivity, impartiality and accountability. However, these are notoriously hard to ensure, as individuals perceive the world subjectively through the prism of their personal mix of our experiences and values. Thus, while the perception of some events may reach a certain degree of consensus (e.g. ongoing natural disaster), others are violently disagreed upon (e.g. claims made about candidates' past in an election campaign). The tension between fact reporting and value- and emotion-based interpretation is also visible in the variety of journalist roles, with reporter and fact checker at one end of the spectrum focusing on facts, and analysts, pundits and columnists at the other end, each interpreting and distilling such reports into a narrative their audience appreciates them for.

The Internet has also changed the landscape in an unprecedented way. Paradoxically, in an era where information has never been so widely available, access to relevant, trustworthy and accurate information is hard to get. In some (mostly online) media sphere, facts widely supported with science are being ignored or marginalized in favor of narratives supported by influence groups; human impact on climate change is the foremost example. Skepticism towards journalists or more generally media has always existed; the Internet is making it easier to lead a social life from which contradiction is absent ("bubble effect") and to be exposed only to content one already agrees with ("echo chamber") and piercing such bubbles with facts is becoming increasingly hard. Also, the multiplication of online media outlets and the possibility to automate misinformation spreading through social bots puts this task out of the reach of human journalists; computerized tools are called for, and are increasingly explored by researchers from many scientific communities.

Since 2015, we have been collaborating with journalists from Le Monde, France's leading national newspaper, in an R&D project named ContentCheck, focused on content management technologies for journalistic fact checking [38]. The goal of this paper is to *analyze and understand what role content management technologies*, seen as data and knowledge management, information extraction, natural language processing, and their interactions, *can play toward facilitating, speeding up and improving fact-checking work*, whether performed by journalists or other fact-checkers.[1] We stress that we do not believe it is feasible to "automate" fact-checking or journalistic work; nor is it desirable, since journalistic content is authored by humans who understand, shape and drive the interests and knowledge of their fellow humans. Instead, we are interested in the role content management can play as a provider of models, algorithms and tools.

Below, we characterize the perimeter of fact checking-work we are interested in, delimit it from other scientific and engineering areas, and highlight its complexity (Section 2); we survey existing research in the realm of fact checking, or related to it, in recent years (Section 3); finally, we identify some open problems, promising research directions, and requirements requirements for systems and platforms that will be developed in the area, still within a content management perspective.

2 COMPUTATIONAL FACT CHECKING

We define the scientific area we consider (Section 2.1), analyze its ingredients (Section 2.2) and finish our outline by acknowledging some of its limitations (Section 2.3).

2.1 Definition and perimeter

Originally, fact checking designated the correctness checking of all facts in news article *before* they are published; this is a core part of the journalistic work. Increasingly though, the term refers to

[1] While most fact-checkers are journalists by trade, others are concerned citizens focusing on a special topic, e.g. water pollution in a given city, scientists working on a topic where politics hotly disputes scientific truth etc.

the analyses of claims *a posteriori*, after a certain article (or tweet, or speech etc.) is published, and often by people unaffiliated with the authors of the original content. Such fact checking work is performed by NGOs maintaining active, high-profile Web sites, such as FactCheck or PolitiFact in the US and FullFact in the UK, by specialists within established news organizations, such as the Fact Checker of the Washington Post[2] in the US, the *Décodeurs*[3] with Le Monde and *Désintox*[4] within Libération, both major newspapers in France, and many others.

Definition: This paper focuses on *a posteriori* fact checking, which we define as the investigative process consisting in:

(1) *extracting claims* from some discourse,
(2) *searching for the facts* the claims are based on,
(3) *assessing the accuracy* of the claim with regards to those backing facts, and
(4) *providing perspective* to claims for which there is no straightforward settlement.

Other definitions and design principles proposed in prior work include [3, 14, 63]. We believe our definition is both general and precise enough to encompass specialized systems capable of solving one of the above tasks, as well as end-to-end systems, providing solutions to several tasks. Broadly speaking, a full-fledged fact checking system enables the construction of specialized knowledge bases, lifting ambiguities inherent in natural language, correcting honest mistakes as well as purposeful misinformation, and archiving the result of such analysis in persistent, searchable, open repositories.

Note that some related scientific areas are left out of the focus of the present work. First, *audio, image and video processing* have been very dynamic on this subject, notably through the field of multimedia forensics (image [6, 57, 70] or video [42, 55]), leading to verification systems and services such as RevEye, Tineye or InVID [61]; we do not cover their specific techniques and instead point to the above references. Second, the area of *fake news detection* is closely related to fact checking, yet we view it as slightly distinct: the former includes any method for classifying a news item as true or false, including methods that do not consider the actual *content* of the news but instead focus on aspects such as its linguistic style or its pattern of propagation in a social network [49, 50, 54]. We consider a technique falls in the scope of fact checking as soon as it examines the *facts (content)* of a claim, usually against some *reference* information. Clearly, fact checking can be seen a worthy component of fake news detection, to which it can bring the distinctive advantage of an *explainable (transparent)* analysis, based on the reference facts.

2.2 Fact checking ingredients

From the above definition, we can derive a set of steps involved in a fact check, up to the possible successful outcome (the fact is confirmed). This is the case if: (i) the fact leaves little to no room for alternative interpretations, (ii) it is backed by sufficient references to sources, (iii) the sources are reliable[5], and (iv) the claim is consistent with the sources.

If the claims is too vague, leading to *too many distinct interpretations*, it is hard or impossible to check it effectively. If the fact checker *lacks access to sufficient reliable sources*, there is not enough background against which to check. If the *reliability* of the sources is in question, the fact check conclusion is hard to trust. Even in the presence of sufficient and reliable sources, evaluating claim accuracy is tricky, due in part to the need to *contextualize* the claim, i.e., understand the precise setting (e.g., in which country, city, at which time) the claim is made; its validity can only be assessed in a specific context. Some claims may be crafted to mislead, i.e., look valid given some context or source that is either irrelevant or flawed.

For any step in the process not to be perceived as biased, *transparency* is key. Indeed, this is apparent in the principles instated by the International Fact-Checking Network (IFCN), an organization sponsored by the Poynter Institute to "promote excellence in fact checking":[6]

(1) A commitment to non-partisanship and fairness.
(2) A commitment to the transparency of sources.
(3) A commitment to the transparency of function and organization.
(4) A commitment to the transparency of methodology.
(5) A commitment to open and honest corrections.

Hence, whichever aspect of the fact checking method research tackles, transparency appears as a horizontal feature that *must* taken into account, for the fact checking output to be useful as opposed to adding more noise to an already noisy landscape.

2.3 The limits of fact checking

While technology can make fact checking easier and faster, its effectiveness is challenged by psychological or cultural barriers. "Confirmation bias" is the well-known psychological observation that people are more likely to believe what fits their prior views, leading to the man-made part of the "echo chamber"; automated recommendation system do the rest, shutting users in their "filter bubbles". The exact extent to which filter bubbles and echo chambers contribute to promoting inaccurate or biased information is still being studied, e.g. [23, 24]. A study [41] has found that defiance towards fact checkers may in fact reinforce a reader's perception if confronted directly (the so-called "backfire effect"), although more recent work on the topic show this may not be a primary matter of concern [65].

Another important issue is the time of the fact check versus the time of the original statement. A compelling article [58] argues that "fact checking is *after-the-fact* checking" in the sense that the verification may come too late, after rumors and false information have had time to "stick" in their audience's minds; the proposed solution is for correct, transparent and verifiable information to reach the audience first. Close to that idea, real-time fact checking aims at reducing the time between publication of the claim and that of their fact checks. However, near instant-correction can exacerbate

[2] https://www.washingtonpost.com/news/fact-checker
[3] http://www.lemonde.fr/les-decodeurs/
[4] http://www.liberation.fr/desintox

[5] There can be disagreement on what is a reliable source. In this paper, we consider a source reliable to the extent that the fact checker trusts it, and/or there is significant consensus (e.g. by a very large majority) on its trustworthiness.
[6] https://www.poynter.org/international-fact-checking-network-fact-checkers-code-principles

the backfire effect [25], ultimately making readers more immune to such corrections. It has also been shown [53] that emotionally engaging information, including rumors and misinformations, spread fast on social networks, which is not always the case for corrections.

Yet, there is also evidence that sufficient and consistent corrections of a false perception may succeed in uprooting it [47]. In a close vein, showing online readers links to "related" stories, include some that correct a misinforming claim on the original page, has been shown to significantly reduce misperceptions [7]; this idea has been very recently taken up by Facebook in order to fight misinformation[7].

The "arms race" between misinformation and fact checking is still on and probably will be so for a while. The focus of this paper is on content management techniques which have been or could be developed toward making fact checking scalable, effective, efficient and transparent.

3 STATE OF THE ART

Figure 1 presents an overview of the tasks involved in fact checking (shown in black boxes), together with their inputs and outputs (shown in blue); main relevant works from the literature appear in their respective tasks.

The central task is to assess the accuracy of a claim accuracy, based on reference sources; this takes as input the claim, and outputs a fact check result or analysis. The claim may have to be extracted from a text source, made available through some media, such as newspapers, social media, political or government communication etc. Reference source search may be needed to identify the reference sources most suited in order to check a given claim. An active area of fact checking work is concerned with putting claims into perspective by analyzing how claim validity is impacted by a slight change in the claim statement. Finally, content management techniques are also called upon to facilitate publishing and sharing fact checking outputs.

The remainder of this section presents the results and techniques involved in each task. For completeness, we start by referring to works defining the fact checking perimeter (Section 3.1); subsequently, we consider: reference data source search (Section 3.2), reference data source construction, integration and refinement (Section 3.3), claim extraction from text (Section 3.4), claim accuracy assessment (Section 3.5), context-based claim interpretation and analysis (Section 3.6) and, last but not least, publishing and sharing fact-checking outputs (Section 3.7).

Few platforms available today implement all (or most) tasks. ClaimBuster [31, 32] (whose functionalities we describe below) is quite complete; a recent vision paper [48] points toward such a comparable architecture, and FullFact also states they are working to develop a complete platform [3]. The CJ Workbench [59] is another example of grass-root initiative to automate journalistic work. This is a very active research and development area today.

3.1 Defining computational fact checking

This very concept is still quite new, and its meaning is being gradually clarified, in an area (going between computer science and journalism) ripe with buzzwords such as fake news, post-truth and the like. We list below advances towards formally defining fact checking.

A pioneering paper on computational journalism [12] did not explicitly mention fact checking but identified many content management functionalities which journalists are advised to rely on, such as information extraction, document exploration, and data integration. Soon after, [14] considered ways in which technology could be used to further journalism values and goals, stated as: truth; public interest; and information. He identified NLP, data mining, machine learning, knowledge representation and information retrieval as computer science areas with the biggest contributions to computational journalism. From an NLP perspective, [63] defines fact checking as the *assignment of a truth value to a claim made in a particular context, on an ordinal scale*, borrowed from the popular Truth-o-Meter scale used by fact checking web sites Politifact ("True", "Mostly True", "Mostly False", etc). It also presents a "golden standard" (fact checks) dataset of manually checked statements, against which automated checking methods can be evaluated.

The recent white-paper by FullFact [3] defines fact checking as a four-stage process where (i) media sources are *monitored*, (ii) claims are *spotted*, (iii) claims are *checked*, (iv) fact checking analysis *results* are created and *published*. This definition is quite consensual among the works we discuss below; claim spotting (also termed claim extraction) and claim accuracy checking have received most significant attention.

3.2 Reference data source search

To check a given claim, one needs to identify suitable reference sources against which to check, and if sources are many, select the most relevant ones.

In FactMinder [27], the authors envision fact-checking as a manual analysis of an article (not a specific claim) published online, against a background of reference data sources. They focus on identifying, in a reference database of RDF knowledge bases and structured documents, those semantically related to the entities (people, places etc.) present in the article to be fact-checked. The technical solution is based on (i) identifying (matching) the article entities to those in the reference RDF graph and (ii) evaluated pre-defined queries to retrieve the content "characterizing" entities of each known type (e.g., political appointments for a politician, albums and songs for a singer etc.). Query results are then displayed to the fact-checker as a background meant to help the check. Such automated background information search is not tailored to a specific claim to check.

DeFacto [26, 34] leverages search engines to collect pages potentially harboring evidence proving or disproving an input claim, considered to be an RDF triple. NLP techniques are applied to translate (enhance) the RDF claim into a set of topic terms which are then looked for in Web pages.

Stance detection is a recent task of information extraction, aiming at determining from a text whether it is in favor of a given target, against it, neutral or unrelated. For example, the tweet *"@realDonaldTrump is the only honest voice of the @GOP"* expresses a positive stance towards the target Donald Trump [2].

[7] https://newsroom.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/

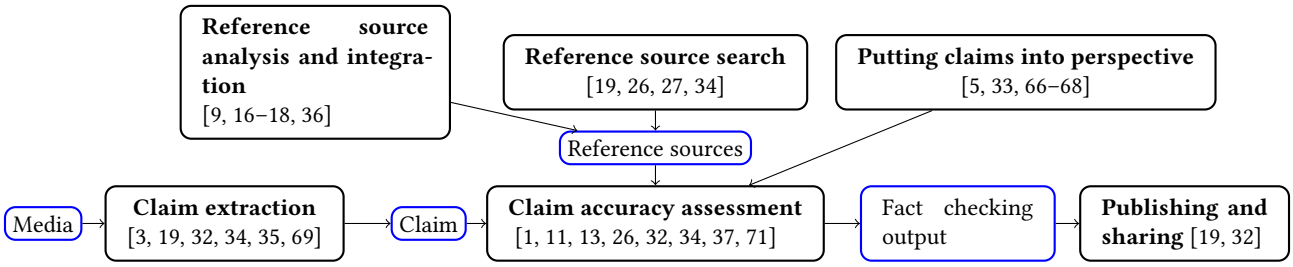

Figure 1: Fact checking tasks, ingredients, and relevant works: an overview.

The textual sources for stance detection can be general claims [4, 35], debates in online forums [30, 40, 56] or student essays [20], but most work have focused on politically- or journalistically-related sources, such as congressional debates [62], tweets [46] (mainly through the organization of a SemEval evaluation campaign in 2016 [39]) or news articles [21] (the Fake News Challenge proposed in 2017 a stance detection task as a first step toward fact-checking[8]).

Last but not least, fact checking outputs are valuable reference sources for future checks. In 2013, Washington Post's TruthTeller project[9] used claims manually checked by Factcheck.org; the Dé-codex plug-in developed by Le Monde also leverages their past fact checking analyses. ClaimBuster [32] relies on a large set of claims manually checked through a crowdsourced effort.

3.3 Reference data source analysis, integration, refinement

The availability, coverage and quality of reference data sources is crucial to the success of fact checking. We mention here areas of research work whose techniques may be beneficially used for fact checking, even though they are applicable in a wider context.

Truth discovery [17, 36] is a task by which the *truthfulness* of facts typically found of online data sources, is assessed based on the *trustworthiness* of those sources. Depending on the approach, both measures can influence one another, i.e., the more truthful facts a site published, the more trustworthy it is, and vice versa. Such large knowledge bases extracted from the Web appear to be good news for journalism, as they provide more reference sources. However, additional verifications are necessary due to their lack of reliability, to inherent errors introduced with automated extraction, to their becoming outdated, etc.

Fact checkers have today at their disposal a wealth of data sources, which for simplicity and ease of use should ideally be exploited together through a single point of access. Data integration, a well-established line of data management research, aims at allowing users to interact with a set of data sources as if it was a single database. While several flavors of data integration systems have been proposed, data spaces [22] appear well-adapted to handling heterogeneous reference sources, independently developed, and available in disparate formats (relational data, HTML or Excel tables, CSV, RDF, JSON frequently found in open data…). Tatooine [8] is a dataspace-style data integration prototype integrating querying, classification, and visualization components for data journalism and fact-checking applications. CJ Workbench [59] is a workflow-oriented platform for retrieving, filtering, storing and analyzing data; it emphasizes collaboration, reuse, and continuously updated views over the data.

3.4 Claim extraction

Claim extraction (also called claim detection) leverages NLP technologies to extract from text, sharp (structured) claims which one may subsequently check for accuracy.

An early attempt at this task in DisputeFinder [19] focused on identifying, in an online text, claims potentially covered by a database of reference containing *disputed claims* (on which contrary opinions have been stated). Phrases are transformed in bags of words through stop word removal, stemming etc. and matched against the database, both in positive mode (to see if the database backs it) and in negative mode (if the contrary is stated in the database). DisputeFinder then shows users additional information on to how or why the claim is disputed, such as related articles or argumentation graph. DisputeFinder also allowed users to add claims the believed were disputed, however a user study showed users sometimes confusing "disputed" with "false", or improperly reported claims and/or sources. This highlights that user expertise is needed in order to contribute meaningfully.

In [35] a formal definition of the task of context-dependent claim detection is proposed: given a topic (short sentence) and a set of relevant articles, their goal is to automatically extract concise statements from the articles that directly support or contests the given topic. The authors also provide manually labeled benchmark dataset and a classification approach based on a cascade of classifiers.

Claim extraction from text feeds (tweeter, newspapers, legislative proceedings), and ultimately, audio and video streams is reported to be also a focus of FullFact ongoing work [3]; in particular, they are interested in automatically identifying the claims covered by prior FullFact manual fact checking work. Technical details are not available at this time.

A follow-up on DeFacto [26] also allows claims to be given in text; in this case, entity disambiguation is performed on the claim, using a reference knowledge base, to bring it to the RDF triple form DeFacto mostly considers.

A line of work closely related to claim extraction is identifying in a text, claims that are worth checking. Machine learning has typically been used for this task. For instance, [43] presents a corpus where claims have been manually classified as verifiable and unverifiable, and [29] uses different kinds of neural networks to learn a classification model. Support Vector Machines is used in

[8] http://www.fakenewschallenge.org/
[9] http://truthteller.washingtonpost.com/

ClaimBuster [31, 32] which monitors data sources such as social media, TV programs and websites, analyses the incoming stream of information and classifies claims in three categories: non-factual (e.g., opinions or subjective content); factual but not interesting (consensual, general); factual and interesting (that is, check-worthy). To train the classifier, a database of 20,000 claims annotated with these categories has been built through crowdsourcing.

Other work try to classify claims as either facts or opinions, [69] from the perspective of an opinion question answering system, [64] for research on deliberation and debate. The distinction is then made between an argument based on facts, from one appealing to feelings and emotions.

An important problem to solve on the way to mostly automated fact-checking is understanding the context in which a given claim is made, that is: identifying the time to which the claim refers, possibly the location or geographical scope, identifying the speaker etc. Indeed, a claim such as "This city's taxes have gone up 20% since the last elections" cannot be checked without such context. These issues are also discussed in [3].

3.5 Claim accuracy assessment

The accuracy of a claim could be simply decided as a value between true and false, or (more frequently) from a multiple-values scale; also, the checking process often produces an analysis or detailed trace, which helps understand the accuracy judgment.

In DeFacto [26, 34], evidence potentially proving the claim is a text snippet found in a Web page, sufficiently close to the claim; fuzzy matching based on similar-term expansion is used to estimate this proximity. The system computes both such a support score (to which extent the proof agrees with the claim) and a trustworthiness score of the Web page comprising the proof. Trustworthiness here is understood as relevance, i.e., the focus is not on establishing a reputation for the page regardless of the query. Subsequently, a supervised machine learning (SVM) classifier is used to decide when there is sufficient evidence to consider that the fact checks. This system has a particular focus on exploiting multilingual evidence.

In the TruthTeller system previously mentioned, a natural language claim extracted from video or audio transcripts with the help of speech recognition, is compared to the previously checked ones from the reference source; a modified version of the Rabin-Karp algorithm using the Levenshtein distance is used.

In [11], authors consider claims as RDF triples, the reference source a knowledge graph such as DBPedia, and cast the task of assessing claim accuracy as a problem of finding short paths in the knowledge graph that connect the claim subject to its object. A truth (or support) value is assigned to each such path, taking into account not only the graph proximity but also the generality of the entities encountered along the path. Here, generality is defined as the number of statements (triples) to which the entity participates: a path is less significant if it passes through very general entities (e.g., Harry Potter and Napoleon are both of Male gender, Male is very general) than if it passes through a very specific one (e.g., Harry Potter and Hermione Granger both study at Hogwarts). Finally, the support of a claim is the highest support value among all paths that support that claim. A more recent study [52] also considers fact checking as a variant of link prediction, and presents a model to assess the meaning and veracity on claims under scrutiny.

ClaimBuster [31, 32] attempts to match the claims against previously checked claims from trusted repositories. If no identical or similar claim can be found, Web search engines and question answering systems such as Wolfram Alpha are leveraged to obtain evidence backing or contradicting the claim. A final step is to combine the responses obtained from this evidence gathering module into a visual interface. In a separate study, the authors show a strong correlation between the output of their system, and that claims checked by professionals at CNN and PolitiFact.

The Fast and Furious FactCheck Challenge[10] proposed to classify news articles (not claims) among the categories True, False, Somewhat True and Somewhat False using any combination of human and automated tools; the 2016 winner states he has used NLP and other AI technologies but not much more is known.

We end this section discussing works which do not meet exactly our fact-checking definition yet come very close to it:

Les Décodeurs, the fact checking team of Le Monde[11] have developed (and share as open data) a database of manual fact checks[12], comprising for each claim a set of Web and social media sources having propagated it, the fact checking analysis, and the final level-of-truth classification. Next, they have developed and share in open source Décodex, a plug-in for navigators and social media like Facebook, which signals to users visiting an information source (a Web page or Facebook account) having published a checked claim, a trust score resulting from the aggregated outputs of previous fact checks over that source. Unlike the journalists who devised it, Décodex does not check fact accuracy, strictly speaking. However, its ability to rate trustworthiness makes it quite relevant to this task.

Last but not least, a closely related (and well-established) field of natural language processing is textual entailment [13], which considers the task of comparing two portions of text and deciding whether the information contained in the first one can be implied from the second. Textual entailment has never been applied explicitly to fact checking problems, but they obviously meet at some points [37]. Many evaluation campaigns and benchmarks are related to textual entailment, as well as paraphrase detection in general, among which PASCAL challenge [13], Answer Validation Exercise [71], the MSRP paraphrase corpus [15] or the SNLI corpus [10]. Most of these tasks and data represent similarity between pairs of text as a binary yes/no classification decision.

The SemEval's Semantic Textual Similarity task [1] offers a graded and typed definition of semantic similarity, which is closer to what fact checking needs, but still focused on general information extraction or machine translation.

Some works also focus on assessing the *credibility* of claims [44, 45], in which case the reputation of sources with matching some topic and stance is evaluated rather than the claims themselves.

3.6 Putting claims into perspective

Even if a claim was found to be true, it may be insignificant or misleading. A typical example introduced in [66] is: a trend (say, a

[10] https://herox.com/factcheck/
[11] http://www.lemonde.fr/les-decodeurs/
[12] http://s1.lemde.fr/mmpub/data/decodex/hoax/hoax_debunks.json

crime rate reduction) which truly holds over a certain interval of time, but had started long before this interval, cannot be attributed to an event at the beginning of the interval (say, someone's election). In this example, the claim significance can be understood framing it into the perspective of a longer time interval. This idea can be generalized, e.g., when the claim uses some aggregated measure at a given granularity in a dimension: a useful perspective may be obtained considering other granularities/dimensions. General statistical and data analysis tools ranging from spreadsheets to OLAP can be used here. More generally, any element related to a claim can be used to assess the claim significance. This opens up the question of a space in which to search for such elements and efficient algorithms and pruning strategies to select such perspectives.

The pioneering work of [66] continued in [67, 68] proposes a framework enabling to formally define the notions of counter-argument (that weakens the original claim), reverse-engineering of vague claims, and claim quality. In this setting, a claim is represented as a parametrized SQL query over some trusted (relational) data sources. The search space is defined by "perturbing" the different query parameters. In addition to general meta-algorithms, the authors focus on comparisons of window aggregate claims and of time series similarity claims, for which they propose specific and efficient algorithms. Windows aggregate claims compare aggregated values computed over two different windows (for example total number of adoptions during two periods of the same duration 1996-2001 and 1990-1995). In this case, perturbations concern the size of windows and their starting time. An example of time series similarity claim is "Person A voted the same as person B $x\%$ of the time during a given period". It requires a similarity function to compare two time series. Perturbations mainly concern the interval.

Data mining techniques are used in [5] to enlighten claims about behaviors in voting or rating contexts. The proposed solution aims at identifying groups of individuals and situations where their agreement significantly differs from usual. In this case, the search space is defined by varying patterns of the form $\langle situation, group_1, group_2 \rangle$, each item being described using attributes.

It is sometimes difficult to draw a clear border between providing a verdict and providing additional information to assess significance. Some works proposed for the former can be used and generalized for the latter. In [33], the author introduces a model in which any fact or ontological rule can be endowed with contextual annotation, stating for instance which sources a certain fact comes from, or whether an axiom holds within a given country or period in time. Coupling thus *hard axioms* with *soft* weighted axioms, a probability is assigned to each context, reflecting what holds in it and whether it entails a contradiction. In this model, claims are represented as conjunctive queries. For each answer to a query, a function maps each context to a score. This enables the user to witness variations in claim credibility w.r.t. to all possible contexts, rather than being providing a boolean, discrete or scalar value to assess its truth.

For claims pertaining to complex issues, system providing a more general and balanced picture of the issue can play an important role. [51] describes a debating system, which takes a claim as input (e.g. "casinos increase criminality"), crawls from a textual corpus in search for article presentation positive and negative stance to the claim, and outputs a summary of the articles for each stances.

3.7 Sharing and publishing fact checking results

An important issue is that of publishing and sharing fact checking outputs in a format that can be computationally used by other tools.

DeFacto [34] envisioned sharing fact checking outputs as RDF graphs endowed with RDF provenance information, specifying which tool has checked the claim and when.

ClaimReview[13] is an open community standard created by Duke in collaboration with Google and Schema.org. It enables fact checkers to annotate web content using a common vocabulary for more effective consumption. For instance, Google uses the annotations to display such fact checks in search results next to the links that are part of the normal Google search results. Technically, Schema.org is a collaborative effort to create, maintain, and promote schemas for exchanging structured data. Content organized according to a Schema.org vocabulary can be encoded in many formats, including JSON.LD (JSON linked data), microdata formats etc. The ClaimReview vocabulary comprises a claim which is of type text, and a claim review, which specializes the pre-existing Review concept from Schema.org. Further, a ClaimReview includes a review rating, and is about a CreativeWork, which can be for instance an Article, Blog, Movie etc. The integration of ClaimReview into the Schema.org vocabulary allows for interoperability with other structured (typically open) data sources.

A database-oriented perspective is taken by the Structured Journalism[14] initiative, which encourages journalists to publish database items, that is, atomized pieces of structured information. Clearly, this can apply to raw data, used as input in the fact checking process, but also to its outputs. Publishing content in such a structured format eases data reuse and allows building several stories by selecting different subsets of the data and focusing on different trends.

As already pointed out, several human-powered fact checking platforms such as FactCheck.org and PolitFact already provide API access, and their output is already used by several other tools. Among automated systems ClaimBuster mentioned above, also provides access to their fact checking outputs. CJWorkbench, already mentioned, claims to enable the sharing of prior work for reuse in similar settings, for instance adapting a voter fraud analysis done on a local election in Texas to a different election in California.

4 PERSPECTIVES

Having completed our study of existing systems, tools and methods, we consider the future. Section 4.1 outlines open issues to be addressed for fact checking, in particular automated techniques, to gain in efficiency and effectiveness. Section 4.2 discusses features of an "ideal" fact check management system, a (so far) inexistent system, specifically designed to support fact checking work.

4.1 Needs and open problems

Progress in several areas is needed to facilitate and empower the development of automated fact checking tools. On the research side, the following steps could be taken:

[13] https://www.blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/
[14] https://reporterslab.org/structured-journalism/

Develop pluridisciplinarity. For several years, interactions between computer scientists and journalists have been extremely fruitful for both sides [14]: discussing with computer scientists can give journalists access to powerful tools for discovering and exposing the truth[15]. This principle is also stated in a recent white-paper from the American Press Institute [28]Social and cognitive sciences also have valuable contributions to bring in order to help devise psychologically effective fact-checking tools. This is not a one-way lane, though: computer science research, and in particular content management, stands to benefit greatly from innovative applications with a strong social impact.

Establish theoretical foundations. To the best of our knowledge, there have not yet been any attempts at establishing theoretical foundations for the computational fact checking task. Such foundations would be a valuable yardstick to validate the well-formedness of approaches, evaluate their coverage and efficiency, and compare their capabilities. Such foundations should provide widely accepted formal definitions for the (canonical) problems in the area, complexity results on those problems, means to solve the most complex ones with approximation guarantees, etc.

Improve transparency. Transparency is a key issue to reach trust. A fact check result should provide answers to common questions such as the reasons for the choice of the claim, who are the fact checkers, the list of used sources with means to evaluate their relevance and credibility, the methodology and tools used. Ideally, the fact check should be reproducible and made available by a neutral entity where everyone can share, analyze, comment, etc.

Ensure explainability. The automatic approaches used in the process of fact checking can be many and complex. They will use many sources, possibly in different languages and complicated structured formats; they will also trigger different kinds of complex algorithms, such as machine learning techniques, most of which building numerical models that behave as black boxes. This raises the question of explainability and interpretability of the results: any system aiming at supporting the journalist or the citizen to make a decision should have an "explain facility".

Develop collaborative tools. Collaboration may range from simple exchanges of trusted data or previous fact checks to coordinated work to face difficult investigations. Furthermore, when fact checkers are of different sensibilities, result credibility improves because it is more difficult to think that a claim has been chosen and treated in a partisan way. CrossCheck[16] is such a collaborative fact checking project; collaboration empowered by content management tools is a strong trend in journalism as large, promoted by organizations such as the ICIJ[17].

Standardize. The introduction of standard fact check formats such as ClaimReview (Section 3.7) is a first step toward standardization, however more standards are needed to cover additional elements (protocols, tools…) used all along the fact checking process (data cleaning, data integration, computations etc.)

[15] A striking recent example is an analysis of the way Wisconsin voting districts are drawn (https://www.nytimes.com/2017/10/06/opinion/sunday/computers-gerrymandering-wisconsin.html), highlighting the (very) low probability that they may result from an "honest" design. In the article's words, "it's math versus math, with democracy at stake".

[16] https://crosscheck.firstdraftnews.com/

[17] International Consortium of Investigative Journalists, behind the Panama Papers and other such high-profile international investigations.

Other steps towards improving the usefulness and effectiveness of fact checking, automated or otherwise, concern journalists and/or the society at large; technology, such as content management tools, could contribute solutions. The most significant ones include:

(a) **Adapt the delivery of fact checking results.** The problem has several linked facets. One is time. On the one hand, it is sometimes important to quickly react to attempt to stop viral dissemination of misinformation. On the other hand, it is also important to take time to provide sharper arguments. The good balance has to be found. A second facet is formulation of the result: a frontal attack on one's convictions and beliefs is not likely to convince him or her. These were mentioned in Section 2.3, when discussing the limits on the effectiveness of manual fact checking; any automated technique must address the same issues. A third facet is the choice of the best media for fact-checking to reach each group of audience.

(b) **Focus more on issues than on claims.** An individual claim may be true or false, but newsworthy questions are usually broader than just a claim. For instance, a misleading statement about the criminal activity of refugees in the countries receiving them, participates to a larger discussion about immigration and the way different political parties argue it should be handled. Thus, effective fact checking should focus more broadly on issues, as noted also in [28]. Existing work on putting claims into perspective (Section 3.6) is a first step in this direction.

(c) **Engage and entertain the audience.** Fact checking success is (also) judged by the audience it can gather and retain. Journalists' talent in making stories out of data-backed proofs needs to be deployed in novel ways, e.g. in "live" mode during political events; this requires near-instant answer from reference source search and claim accuracy checking tools. Journalism at large (and fact checking in particular) may also appeal to users through innovative Web-based media such as interactive, structured data exploration components, embedded inside Web pages. Fact checking components designed directly as plug-ins for social media are another example of journalism moving where the audience is; content management tools are typical tools leveraged in the development of such plug-ins.

(d) **Educate.** Just as computer literacy is gaining ground through its presence in school curricula, data literacy, envisioned as a set of math and statistic skills, can spread through dedicated education modules at all levels[18]; understanding the way media and communication works can give the public further tools to discern manipulation, statistic or otherwise.

4.2 Towards a fact check management system

From our identified fact checking tasks, the analysis of existing works and techniques, as well as the needs and desiderata outlined above, we can derive the functionalities of an "ideal" fact check management system (FCMS), that is, a content management system (CMS) tailored to the needs and uses that fact checking entails:

Data storage. The FCMS should be capable of storing data sources in various formats, such as they come in journalists' hands:

[18] See e.g. the course "Calling Bullshit: Data Reasoning in a Digital World" created at U. Washington, http://callingbullshit.org/syllabus.html

PDF, structured text, JSON, CSV, relational databases etc. Automatically detecting the data type and relying on an appropriate system to store each kind of data (or simply a file system when nothing more is available) would simplify users' experience. Automatic back-up mechanisms e.g. toward a cloud-storage account would prevent accidental loss of data and allow data reuse. Surprising as it may seem, data used in investigations is not always saved persistently by journalists; an acceptable CMS must make the storage and retrieval experience as smooth as possible.

Data matching, linking and integration. The FCMS should allow users to identify connections between independently produced data sources, so that they can be exploited together. Many techniques developed for data integration and for Linked Data production can be applied here. Data processing functionalities as described in this and the previous item are especially important if the FCMS is to support gathering information and checking/investigating broader issues, as discussed in Section 4.1, point b.

NLP: large and small. As shown throughout this paper, NLP techniques are crucial for working with text, the predominant form of information sharing today. Building highly accurate NLP tools for very specific tasks is and will probably stay out of reach for non specialists for a while. However, basic NLP functionalities such as named entity recognition from a knowledge base or structured database, transforming documents in bag of words, enhancing it through synonyms and close terms etc. are by now well understood and well supported by tools including open-source ones. They need to be made available and easy to integrate within the FCMS.

Time management. Being able to trace the data and its evolution is important for accuracy, transparency, reproducibility. FCMS should be able to record and permanently store various pieces of time information such as: data creation time stamp (immutable publication date of an article or a dataset), acquisition times (can be repeated several times as the data evolves), statement date (when people made a certain statement etc.) Any kind of data is concerned including results of fact check.

Fact and validity time. Facts and events almost all have a limited period of validity. This information can be explicit in the dataset (e.g. a knowledge base gives the starting and ending dates of World War I, or the period where W. Churchill was Prime Minister of the U.K. [60]); but it may need to be inferred from the different snapshots or versions of a same dataset, e.g., a list of Twitter accounts of all ministers is not time-stamped, but evolves with time. In order to understand when a claim is or is not valid, and to ensure not to lose any information through time, all facts should be associated with a validity time, and all query languages and systems should be able to handle it.

Data quality management. This point encompasses all the tools for managing the life cycle of reference sources and fact checks. This is related to provenance and lineage management which enable to trace what data and operations are involved in producing a given result (data, fact check…).

Support for reproducibility. In some way, fact checking may be seen as a scientific or forensic work, for which reproducibility is needed.[19] Ideally, it should be able to record the trace of a fact checking effort so that one could "replay" it on the same inputs with the same tool, and get the same result.

Modularity. Using today's technology, the processes involved in fact-checked are extremely time consuming. The FCMS should be designed from the start with modularity and reuse in mind, so that claim analysis pipelines can be easily built and their components re-used. This will offer a great advantage over "one-off" systems.

Compliance with the standards. The FCMS should adopt all standards of the trade and be open in the export and sharing of its inputs and outputs.

[19] http://ropensci.github.io/reproducibility-guide/

5 CONCLUSION

This work provides a characterization of automated fact checking as a content management problem, drawing from data and knowledge management, natural language processing and information retrieval; machine learning is also leveraged for many of the tasks involved. Based on this characterization, we have identified the main tasks involved, and surveyed existing content management tools and techniques which have and/or could be used to implement the tasks. While we found the area generates a lot of excitement and many works have been put forward, most efforts are currently quite disparate, thus a main goal of our analysis has been to classify, compare, and see how current techniques could fit together. We identified a set of areas where we believe more content management research should be invested to help automated fact checking advance, and described a blueprint architecture for an ideal fact check management system; we view it as a frame for assembling (and thinking about) research efforts currently ongoing in the area.

ACKNOWLEDGMENTS

This work is partially supported by the French National Research Agency grant ANR-15-CE23-0025-01 (ContentCheck project) and by the KAKENHI grant number 17K12786, together with the New Energy and Industrial Technology Development Organization (NEDO), Japan.

REFERENCES

[1] Eneko Agirre, Daniel Cer, Mona Diab, and Aitor Gonzalez-Agirre. 2013. *Sem 2013 Shared Task: Semantic Textual Similarity. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics and the Shared Task: Semantic Textual Similarity*. 32–43.

[2] Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance Detection with Bidirectional Conditional Encoding. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Jian Su, Kevin Duh, and Xavier Carreras (Eds.). Association for Computational Linguistics, Austin, Texas, 876–885. https://aclweb.org/anthology/D16-1084

[3] Mevan Babakar and Will Moy. 2016. The State of Automated Factchecking. https://fullfact.org/media/uploads/full_fact-the_state_of_automated_factchecking_aug_2016.pdf. (2016).

[4] Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. 2017. Stance Classification of Context-Dependent Claims. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Association for Computational Linguistics, Valencia, Spain, 251–261. http://www.aclweb.org/anthology/E17-1024

[5] Adnene Belfodil, Sylvie Cazalens, Philippe Lamarre, and Marc Plantevit. 2017. Flash Points: Discovering Exceptional Pairwise Behaviors in Vote or Rating Data. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18-22, 2017, Proceedings, Part II (Lecture Notes in Computer Science)*, Michelangelo Ceci, Jaakko Hollmén, Ljupco Todorovski, Celine Vens, and Saso Dzeroski (Eds.), Vol. 10535. Springer, 442–458. DOI: https://doi.org/10.1007/978-3-319-71246-8_27

[6] Gajanan K. Birajdar and Vijay H. Mankar. 2013. Digital Image Forgery Detection Using Passive Techniques: A Survey. *Digit. Investig.* 10, 3 (Oct. 2013), 226–245.

DOI : https://doi.org/10.1016/j.diin.2013.04.007

[7] Leticia Bode and Emily K. Vraga. 2015. In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media. *Journal of Communication* 65, 4 (2015), 619–638. DOI : https://doi.org/10.1111/jcom.12166

[8] Raphaël Bonaque, T. D. Cao, Bogdan Cautis, François Goasdoué, J. Letelier, Ioana Manolescu, O. Mendoza, S. Ribeiro, Xavier Tannier, and Michaël Thomazo. 2016. Mixed-instance querying: a lightweight integration architecture for data journalism. *PVLDB* 9, 13 (2016), 1513–1516.

[9] Allan Borodin, Gareth O. Roberts, Jeffrey S. Rosenthal, and Panayiotis Tsaparas. 2005. Link Analysis Ranking: Algorithms, Theory, and Experiments. *ACM Trans. Internet Technol.* 5, 1 (Feb. 2005), 231–297. DOI : https://doi.org/10.1145/1052934.1052942

[10] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.

[11] Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M. Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. 2015. Computational fact checking from knowledge networks. *PloS one* 10, 6 (2015), e0128193.

[12] Sarah Cohen, Hamilton James T., and Fred Turner. 2011. Computational Journalism. *Commun. ACM* 54, 10 (Oct. 2011), 66–71. DOI : https://doi.org/10.1145/2001269.2001288

[13] Ido Dagan, Oren Glickman, and Bernardo Magnini. 2005. The PASCAL Recognising Textual Entailment Challenge. In *PASCAL Challenges Workshop for Recognizing Textual Entailment*. http://oren.glickman.com/publications/LNAI_39440177.pdf

[14] Nicholas Diakopoulos. 2012. Cultivating the Landscape of Innovation in Computational Journalism. http://www.nickdiakopoulos.com/wp-content/uploads/2012/05/diakopoulos_whitepaper_systematicinnovation.pdf. (2012).

[15] B. Dolan, C. Quirk, and C. Brockett. 2004. Unsupervised Construction of Large Paraphrase Corpora: Exploiting Massively Parallel News Sources. In *Proceedings of the 20th International Conference on Computational Linguistics (Coling 04)*. COLING, Geneva, Switzerland.

[16] Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge Vault: A Web-scale Approach to Probabilistic Knowledge Fusion. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*. ACM, New York, NY, USA, 601–610. DOI : https://doi.org/10.1145/2623330.2623623

[17] Xin Luna Dong, Laure Berti-Equille, and Divesh Srivastava. 2009. Truth discovery and copying detection in a dynamic world. *Proceedings of the VLDB Endowment* 2, 1 (2009), 562–573.

[18] Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, and Wei Zhang. 2015. Knowledge-Based Trust: Estimating the Trustworthiness of Web Sources. In *Proceedings of the VLDB Endowment*. http://arxiv.org/pdf/1502.03519v1.pdf

[19] Rob Ennals, Beth Trushkowsky, and John Mark Agosta. 2010. Highlighting disputed claims on the web. In *Proceedings of the 19th international conference on World wide web*. ACM, 341–350.

[20] Adam Faulkner. 2014. Automated Classification of Stance in Student Essays: An Approach Using Stance Target Information and the Wikipedia Link-Based Measure. In *Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference*. Association for the Advancement of Artificial Intelligence, Pensacola Beach, USA. https://pdfs.semanticscholar.org/2e8d/01e2fcf7ad7bfc889360a7c9c495effbdc34.pdf

[21] William Ferreira and Andreas Vlachos. 2016. Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, San Diego, California, 1163–1168. http://www.aclweb.org/anthology/N16-1138

[22] Michael Franklin, Alon Halevy, and David Maier. 2005. From Databases to Dataspaces: A New Abstraction for Information Management. *SIGMOD Record* 34, 4 (Dec. 2005), 27–33. DOI : https://doi.org/10.1145/1107499.1107502

[23] R. Kelly Garrett. 2009. Echo chambers online?: Politically motivated selective exposure among Internet news users1. *Journal of Computer-Mediated Communication* 14, 2 (2009), 265–285. DOI : https://doi.org/10.1111/j.1083-6101.2009.01440.x

[24] R. Kelly Garrett. 2016. Facebook's problem is more complicated than fake news. http://theconversation.com/facebooks-problem-is-more-complicated-than-fake-news-68886. (2016).

[25] R. Kelly Garrett and Brian E. Weeks. 2013. The promise and peril of real-time corrections to political misperceptions. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1047–1058.

[26] Daniel Gerber, Diego Esteves, Jens Lehmann, Lorenz Bühmann, Ricardo Usbeck, Axel-Cyrille Ngonga Ngomo, and René Speck. 2015. DeFacto—Temporal and multilingual Deep Fact Validation. *Web Semantics: Science, Services and Agents on the World Wide Web* 35 (2015), 85–101.

[27] François Goasdoué, Konstantinos Karanasos, Yannis Katsis, Julien Leblay, Ioana Manolescu, and Stamatis Zampetakis. 2013. Fact checking and analyzing the web. In *SIGMOD*, Kenneth A. Ross, Divesh Srivastava, and Dimitris Papadias (Eds.). ACM, 997–1000. DOI : https://doi.org/10.1145/2463676.2463692

[28] Alan Greenblatt. 2017. The Future of Fact-Checking: Moving ahead in political accountability journalism. https://www.americanpressinstitute.org/publications/reports/white-papers/future-of-fact-checking/. (2017).

[29] Chinnappa Guggilla, Tristan Miller, and Iryna Gurevych. 2016. CNN- and LSTM-based Claim Classification in Online User Comments. In *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers (COLING 2016)*. 2740–2751. https://www.ukp.tu-darmstadt.de/fileadmin/user_upload/Group_UKP/publikationen/2016/2016_COLING_CG.pdf

[30] Kazi Saidul Hasan and Vincent Ng. 2013. Stance Classification of Ideological Debates: Data, Models, Features, and Constraints. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. Asian Federation of Natural Language Processing, Nagoya, Japan, 1348–1356. http://www.aclweb.org/anthology/I13-1191

[31] Naeemul Hassan, Chengkai Li, and Mark Tremayne. 2015. Detecting check-worthy factual claims in presidential debates. In *Proceedings of International on Conference on Information and Knowledge Management*. ACM, 1835–1838.

[32] Naeemul Hassan, Gensheng Zhang, Fatma Arslan, Josue Caraballo, Damian Jimenez, Siddhant Gawsane, Shohedul Hasan, Minumol Joseph, Aaditya Kulkarni, Anil Kumar Nayak, and others. 2017. ClaimBuster: The First-ever End-to-end Fact-checking System. *Proceedings of the VLDB Endowment* 10, 7 (2017).

[33] Julien Leblay. 2017. A Declarative Approach to Data-Driven Fact Checking.. In *AAAI*. 147–153.

[34] Jens Lehmann, Daniel Gerber, Mohamed Morsey, and Axel-Cyrille Ngonga Ngomo. 2012. DeFacto-deep fact validation. In *International Semantic Web Conference*. Springer, 312–327.

[35] Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context Dependent Claim Detection. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*. Dublin City University and Association for Computational Linguistics, Dublin, Ireland, 1489–1500. http://www.aclweb.org/anthology/C14-1141

[36] Yaliang Li, Jing Gao, Chuishi Meng, Qi Li, Lu Su, Bo Zhao, Wei Fan, and Jiawei Han. 2016. A survey on truth discovery. *Acm Sigkdd Explorations Newsletter* 17, 2 (2016), 1–16.

[37] Amnon Lotan, Asher Stern, and Ido Dagan. 2013. TruthTeller: Annotating Predicate Truth. In *Proceedings of NAACL-HLT 2013*. Atlanta, USA, 752–757. http://www.aclweb.org/anthology/N13-1091

[38] Ioana Manolescu. 2017. ContentCheck: Content Management Techniques and Tools for Fact-checking. *ERCIM News* (Oct. 2017). https://hal.inria.fr/hal-01596563

[39] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 Task 6: Detecting Stance in Tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, 31–41. http://www.aclweb.org/anthology/S16-1003

[40] Akiko Murakami and Rudy Raymond. 2010. Support or Oppose? Classifying Positions in Online Debates from Reply Activities and Opinion Expressions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*. Beijing, China, 869–875. http://www.aclweb.org/anthology/C10-2100

[41] Brendan Nyhan and Jason Reifler. 2010. When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior* 32, 2 (01 Jun 2010), 303–330. DOI : https://doi.org/10.1007/s11109-010-9112-2

[42] Ramesh C. Pandey, Sanjay K. Singh, and Kaushal K. Shukla. 2016. Passive Forensics in Image and Video Using Noise Features: A Review. *Digital Investigation* 19, C (Dec. 2016), 1–28. DOI : https://doi.org/10.1016/j.diin.2016.08.002

[43] Joonsuk Park and Claire Cardie. 2014. Identifying Appropriate Support for Propositions in Online User Comments. In *Proceedings of the First Workshop on Argumentation Mining*. Association for Computational Linguistics, Baltimore, Maryland, 29–38. http://www.aclweb.org/anthology/W14-2105

[44] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2016. Credibility Assessment of Textual Claims on the Web. In *CIKM*. ACM, 2173–2178.

[45] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2017. Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media. In *Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 1003–1012. DOI : https://doi.org/10.1145/3041021.3055133

[46] Ashwin Rajadesingan and Huan Liu. 2014. Identifying Users with Opposing Opinions in Twitter Debates. In *7th International Conference Social Computing, Behavioral-Cultural Modeling and Prediction*, William G. Kennedy, Nitin Agarwal, and Shanchieh Jay Yang (Eds.). Springer International Publishing, Cham, 153–160. DOI : https://doi.org/10.1007/978-3-319-05579-4_19

[47] David P. Redlawsk, Andrew JW. Civettini, and Karen M. Emmerson. 2010. The affective tipping point: Do motivated reasoners ever "get it"? *Political Psychology*

31, 4 (2010), 563–593.

[48] Victoria L. Rubin. 2017. News Verification Suite: Towards System Design to Supplement Reporters' and Editors' Judgements. In *Proceedings of the 45th Annual Conference of The Canadian Association for Information Science/ L'Association canadienne des sciences de l'information (CAIS/ACSI2017)*.

[49] Victoria L. Rubin, Yimin Chen, and Niall J. Conroy. 2015. Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology* 52, 1 (2015), 1–4.

[50] Victoria L. Rubin, Niall J. Conroy, Yimin Chen, and Sarah Cornwell. 2016. Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News.. In *Proceedings of NAACL-HLT*. 7–17.

[51] Misa Sato, Kohsuke Yanai, Toshinori Miyoshi, Toshihiko Yanase, Makoto Iwayama, Qinghua Sun, and Yoshiki Niwa. 2015. End-to-end Argument Generation System in Debating. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL), System Demonstrations*. 109–114. http://aclweb.org/anthology/P/P15/P15-4019.pdf

[52] Baoxu Shi and Tim Weninger. 2016. Fact Checking in Heterogeneous Information Networks. In *Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 101–102. DOI: https://doi.org/10.1145/2872518.2889354

[53] Jieun Shin, Lian Jian, Kevin Driscoll, and François Bar. 2017. Political rumoring on Twitter during the 2012 US presidential election: Rumor diffusion and correction. *New Media & Society* 19, 8 (2017), 1214–1235. DOI: https://doi.org/10.1177/1461444816634054 arXiv:https://doi.org/10.1177/1461444816634054

[54] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter* 19, 1 (2017), 22–36.

[55] K. Sitara and B.M. Mehtre. 2016. Digital Video Tampering Detection: An Overview of Passive Techniques. *Digital Investigation* 18, C (Sept. 2016), 8–22. DOI: https://doi.org/10.1016/j.diin.2016.06.003

[56] Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing Stances in Online Debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. Association for Computational Linguistics, Suntec, Singapore, 226–234. http://www.aclweb.org/anthology/P/P09/P09-1026

[57] Matthew C. Stamm, Min Wu, and Kuo J. Ray Liu. 2013. Information Forensics: An Overview of the First Decade. *IEEE Access* 1 (2013), 167–200. DOI: https://doi.org/10.1109/ACCESS.2013.2260814

[58] Jonathan Stray. 2017. Defense Against the Dark Arts: Networked Propaganda and Counter-Propaganda. http://jonathanstray.com/networked-propaganda-and-counter-propaganda. (2017).

[59] Jonathan Stray. 2017. Introducing the CJ Workbench. http://jonathanstray.com/introducing-the-cj-workbench. (2017).

[60] Mihai Surdeanu. 2013. Overview of the TAC2013 Knowledge Bbase Population Evaluation: English Slot Filling and Temporal Slot Filling. In *Proceedings of the TAC-KBP 2013 Workshop*.

[61] Denis Teyssou, Jean-Michel Leung, Evlampios Apostolidis, Konstantinos Apostolidis, Symeon Papadopoulos, Markos Zampoglou, Olga Papadopoulou, and Vasileios Mezaris. 2017. The InVID Plug-in: Web Video Verification on the Browser. In *Proceedings of the 1st International Workshop on Multimedia Verification*. Mountain View, USA. http://www.iti.gr/~bmezaris/publications/mm17_2_preprint.pdf

[62] Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Sydney, Australia, 327–335. https://www.cs.cornell.edu/home/llee/papers/tpl-convote.dec06.pdf

[63] Andreas Vlachos and Sebastian Riedel. 2014. Fact Checking: Task definition and dataset construction. In *ACL 2014 Workshop on Language Technologies and Computational Social Science*. 18–22. http://aclweb.org/anthology/W14-2508

[64] Marilyn Walker, Jean Fox Tree, Pranav Anand, Rob Abbott, and Joseph King. 2012. A Corpus for Research on Deliberation and Debate. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)* (23-25), Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association (ELRA), Istanbul, Turkey.

[65] Thomas Wood and Ethan Porter. 2016. The elusive backfire effect: mass attitudes' steadfast factual adherence. *Political Behavior* (2016), 1–29.

[66] You Wu, Pankaj K Agarwal, Chengkai Li, Jun Yang, and Cong Yu. 2014. Toward computational fact-checking. *Proceedings of the VLDB Endowment* 7, 7 (2014), 589–600.

[67] You Wu, Pankaj K Agarwal, Chengkai Li, Jun Yang, and Cong Yu. 2017. Computational Fact Checking through Query Perturbations. *ACM Transactions on Database Systems (TODS)* 42, 1 (2017), 4.

[68] You Wu, Junyang Gao, Pankaj K Agarwal, and Jun Yang. 2017. Finding diverse, high-value representatives on a surface of answers. *Proceedings of the VLDB Endowment* 10, 7 (2017), 793–804.

[69] Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. In *Proceedings of 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 129–136.

[70] Markos Zampoglou, Symeon Papadopoulos, and Yiannis Kompatsiaris. 2017. Large-scale Evaluation of Splicing Localization Algorithms for Web Images. *Multimedia Tools Appl.* 76, 4 (Feb. 2017), 4801–4834. DOI: https://doi.org/10.1007/s11042-016-3795-2

[71] Álvaro Rodrigo, Anselmo Peñas, and Felisa Verdejo. 2008. Overview of the Answer Validation Exercise 2008. In *Proceedings of the 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008*. 296–313.