

2017

News Verification Suite: Towards System Design to Supplement Reporters' and Editors' Judgements

Victoria L. Rubin

Western University, vrubin@uwo.ca

Follow this and additional works at: <https://ir.lib.uwo.ca/fimspres>



Part of the [Library and Information Science Commons](#)

Citation of this paper:

Rubin V. L. (2017). News Verification Suite: Towards System Design to Supplement Reporters' and Editors' Judgements. In the Proceedings of the 45th Annual Conference of The Canadian Association for Information Science/ L'Association canadienne des sciences de l'information (CAIS/ACSI2017), Ryerson University, Toronto, May 31 - June 2, 2017.

CAIS/ACSI 2017

NEWS VERIFICATION SUITE:

TOWARDS A SYSTEM DESIGN TO SUPPLEMENT
REPORTERS' AND EDITORS' JUDGEMENTS

VICTORIA RUBIN



Western
FIMS

Faculty of Information & Media Studies

LIT RL

Language &
i
nformation
T
echnology
Research Lab
VICTORIA RUBIN

GOALS

- *A conceptual design. A vision.*
- *Meant to unlock some of the problems.*

A SET OF TECHNOLOGICAL INTERVENTIONS



*To support news production + consumption
To combat Misinformation | Deception | “Fakes” =*

THE NEWS VERIFICATION SUITE





AGENDA:

Explaining the “**mechanics**” of putting the knowledge from 3 points of view

- 1. what's advisable? (LIS/Education/Media Literacy)*
- 2. what's needed? (Journalism)*
- 3. what's feasible? (NLP/Automation)*

SO-CALLED “FAKE NEWS” PROBLEM

Digital Deception (Hancock, 2012)

= an intentional control of information to create a false belief or false conclusion in digital environments

Several types – several “locks” to pick:

Outright falsifications (Silverman, 2017)

- for propaganda
- for monetary gain



Journalistic fraud, rumors, hoaxes, clickbait (Chen, Conroy, & Rubin, 2015);

Misleading satire (Rubin, Conroy, Chen, & Cornwell, 2016);

Native ads (Cornwell & Rubin, 2017).



UNDERLYING ISSUES & LIS SOLUTIONS

Credibility results from an interaction of **source + message + receiver characteristics** (Wathen, Burkell, 2002).

Quality of information is determined by accuracy, believability, reputation, objectivity (Wang & Strong, 1996), factuality (Zmud, 1978), credibility, consistency, completeness (Jarke & Vassiliou, 1997), correctness, lack of ambiguity (Wang & Wang, 1996), precision, reliability, freedom from bias (DeLone & McLean, 1992).

LIS SOLUTIONS:

check-lists, manual verification, seals of credibility, training in critical thinking, etc.



LIS SOLUTIONS MANUAL



Step-by-Step Check-Lists

HOW TO SPOT FAKE NEWS

- CONSIDER THE SOURCE**
Click away from the story to investigate the site, its mission and its contact info.
- READ BEYOND**
Headlines can be outrageous in an effort to get clicks. What's the whole story?
- CHECK THE AUTHOR**
Do a quick search on the author. Are they credible? Are they real?
- SUPPORTING SOURCES?**
Click on those links. Determine if the info given actually supports the story.
- CHECK THE DATE**
Reposting old news stories doesn't mean they're relevant to current events.
- IS IT A JOKE?**
If it is too outlandish, it might be satire. Research the site and author to be sure.
- CHECK YOUR BIASES**
Consider if your own beliefs could affect your judgement.
- ASK THE EXPERTS**
Ask a librarian, or consult a fact-checking site.

IFLA
International Federation of Library Associations and Institutions



LIS SOLUTIONS

MANUAL



NLP SOLUTIONS

AUTOMATED



Step-by-Step Check-Lists

HOW TO SPOT FAKE NEWS

- CONSIDER THE SOURCE**
Click away from the story to investigate the site, its mission and its contact info.
- READ BEYOND**
Headlines can be outrageous in an effort to get clicks. What's the whole story?
- CHECK THE AUTHOR**
Do a quick search on the author. Are they credible? Are they real?
- SUPPORTING SOURCES?**
Click on those links. Determine if the info given actually supports the story.
- CHECK THE DATE**
Reposting old news stories doesn't mean they're relevant to current events.
- IS IT A JOKE?**
If it is too outlandish, it might be satire. Research the site and author to be sure.
- CHECK YOUR BIASES**
Consider if your own beliefs could affect your judgement.
- ASK THE EXPERTS**
Ask a librarian, or consult a fact-checking site.

IFLA
International Federation of Library Associations and Institutions

Variety of Specialized Algorithms

Automated Deception Detection
63% accuracy (Rubin, et al., 2015).

LIS SOLUTIONS

MANUAL



NLP SOLUTIONS

AUTOMATED



Step-by-Step Check-Lists

HOW TO SPOT FAKE NEWS

- CONSIDER THE SOURCE**
Click away from the story to investigate the site, its mission and its contact info.
- READ BEYOND**
Headlines can be outrageous in an effort to get clicks. What's the whole story?
- CHECK THE AUTHOR**
Do a quick search on the author. Are they credible? Are they real?
- SUPPORTING SOURCES?**
Click on those links. Determine if the info given actually supports the story.
- CHECK THE DATE**
Reposting old news stories doesn't mean they're relevant to current events.
- IS IT A JOKE?**
If it is too outlandish, it might be satire. Research the site and author to be sure.
- CHECK YOUR BIASES**
Consider if your own beliefs could affect your judgement.
- ASK THE EXPERTS**
Ask a librarian, or consult a fact-checking site.

Variety of Specialized Algorithms

Satire Detector: 90% precision & 84% recall
(Rubin et al., 2016).

Copy & paste
an *unattributed*
article:
headlines +
body of text



Mobile or Web

satiredetector.fims.uwo.ca

Satire Detector

Regime By Placing New Collection Of Tees In Drawer

Body of the article: (1000 word maximum)

AUGUSTA, ME—In a dark harbinger of troubled times to come, recently hired office manager Mel Pritchard reportedly placed a new assortment of tea bags in Graydon Media's kitchen drawer Friday, the first foreshadowing signal of what to expect from her regime. "I guess now that Mel's seized the reins of power, we have to somehow get by on nothing but chamomile, chai, and Earl Grey," said marketing

Satire Detector

The article seems to be an example of: **SATIRE**

The Satire Detector is 84% confident of the prediction.

Is this prediction

Incorrect?

Correct?

Unsure?

Based on its text features, it is considered **NOT ABSURD** and **HUMOROUS**

The article has a high occurrence of **Pronouns, Prepositions, Verbs**

Please enter the source URL of the article: (if available)

Thank you for your feedback!



<http://SatireDetector.fims.uwo.ca/>

EXCELLENCE CRITERIA IN REPORTING & LIS

***JOURNALISTS* ask 6 key veracity questions:**

1. Kind of content?
2. Complete?
3. Sources?
4. Evidence? Vetting?
5. Alternative explanations?
6. Am I learning what I need to?

(Kovach & Rosenstiel, 2010, p. 32)



***EDITORS* assess newsworthiness:**

big picture composition, completeness, conflict, consequence, continuity, controversy, elite countries and persons, importance, interest, negativity, newsworthiness, novelty, oddity, predictability, relevance, sensationalism, timeliness, unexpectedness, and veracity

(Galtung & Ruge, 1973; Shoemaker & Mayfield, 1987).

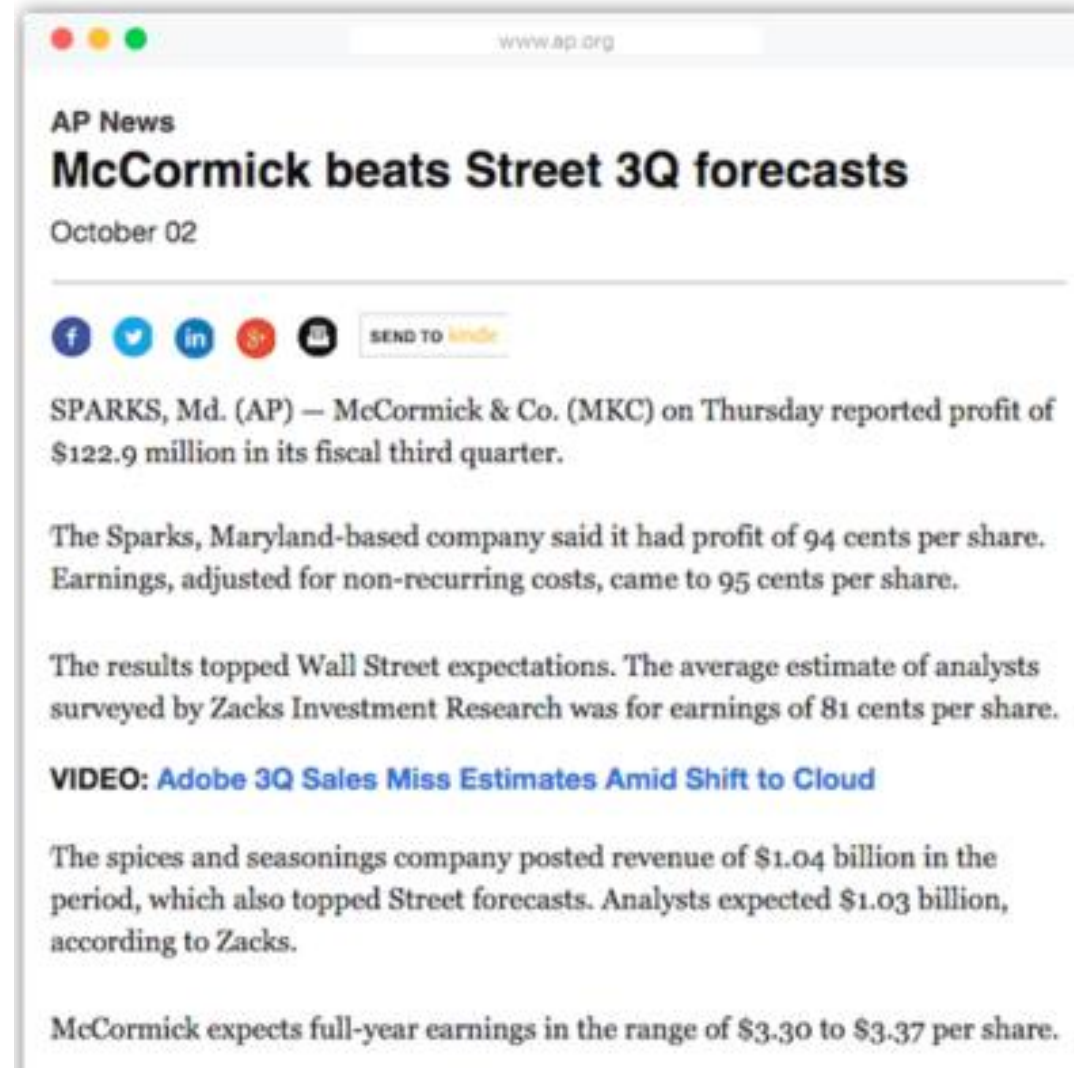
LIS FACTORS FOR HIGH QUALITY OF INFORMATION:

accuracy, believability, reputation, objectivity, factuality, credibility, consistency, completeness, correctness, lack of ambiguity, precision, reliability, lack of bias.

EXAMPLE USE OF NLP IN NEWSROOMS

The Associated Press uses Automated Insight's "Wordsmith" to transform raw earnings data into **3,700 corporate earnings stories every quarter.**

In 2017 "**The Future of Augmented Journalism: A guide for newsrooms in the age of smart machines**" by Marconi, Siegman, & Machine Journalist.



www.ap.org

AP News
McCormick beats Street 3Q forecasts
October 02

SPARKS, Md. (AP) — McCormick & Co. (MKC) on Thursday reported profit of \$122.9 million in its fiscal third quarter.

The Sparks, Maryland-based company said it had profit of 94 cents per share. Earnings, adjusted for non-recurring costs, came to 95 cents per share.

The results topped Wall Street expectations. The average estimate of analysts surveyed by Zacks Investment Research was for earnings of 81 cents per share.

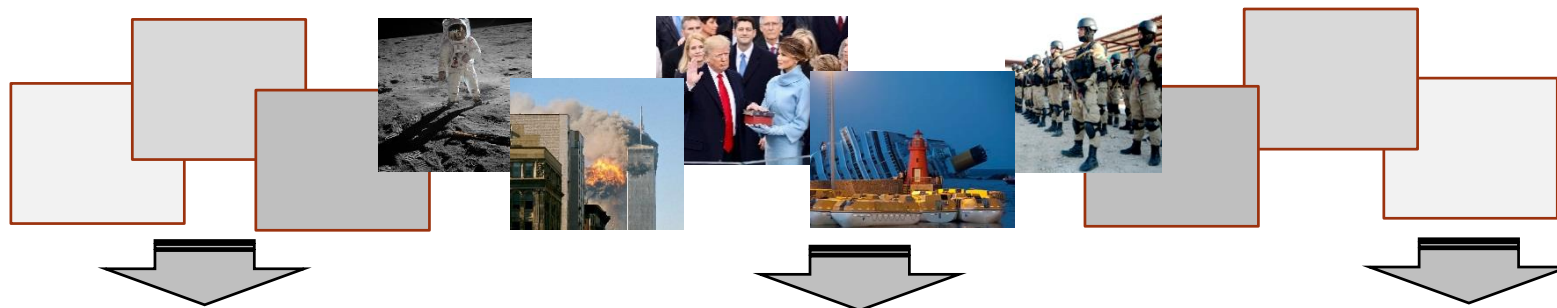
VIDEO: Adobe 3Q Sales Miss Estimates Amid Shift to Cloud

The spices and seasonings company posted revenue of \$1.04 billion in the period, which also topped Street forecasts. Analysts expected \$1.03 billion, according to Zacks.

McCormick expects full-year earnings in the range of \$3.30 to \$3.37 per share.

TRADITIONAL NEWS DISSEMINATION MODEL

Events (World or Local)



Sources (Direct Participants, Eye-Witnesses, Informants, etc.) + Experts + Previous Stories/Reports

Traditional Journalists

Fact-Checkers, Proof-Readers,
Editors, Lawyers

News
Stories

Retractions
Corrections

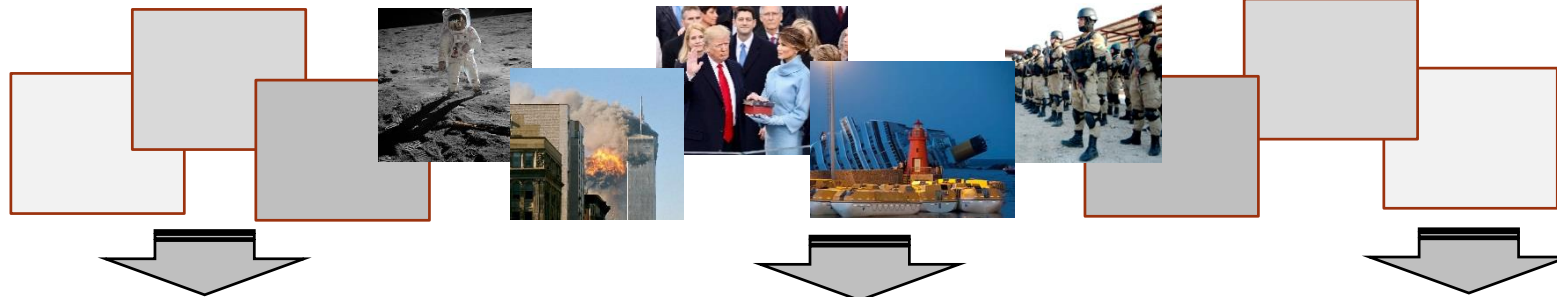
Audience, Public Editors



TRADITIONAL NEWS DISSEMINATION MODEL

REPORTING A.S.A.P. MODEL

Events (World or Local)



Sources (Direct Participants, Eye-Witnesses, Informants, etc.) + Experts + Previous Stories/Reports

Traditional Journalists

Content Creators ("Reporters a.s.a.p.")

Fact-Checkers, Proof-Readers,
Editors, Lawyers

Stories

News
Stories

Retractions
Corrections

Social Media Amplification

Audience, Public Editors

Audience, Public Editors

Mass Accountability

LiT RL

Victoria Rubin. CAIS/ACSI 2017



TRADITIONAL NEWS DISSEMINATION MODEL

REPORTING A.S.A.P. MODEL

TECH. INTERVENTIONS

Events (World or Local)



Opportunities for text analysis to support human decision making

NEWS VERIFICATION SUITE



Sources (Direct Participants, Eye-Witnesses, Informants, etc.) + Experts + Previous Stories/Reports

Traditional Journalists

Content Creators ("Reporters a.s.a.p.")

Fact-Checkers, Proof-Readers, Editors, Lawyers

Stories

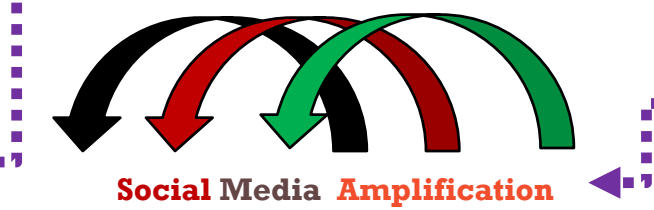
Deception Detection
Topicality Detection
Novelty Detection

Fraud Detection
Sentiment Analysis
Auto-Summarization

Rumor Debunking
Satire Labeling
Fact-Checking

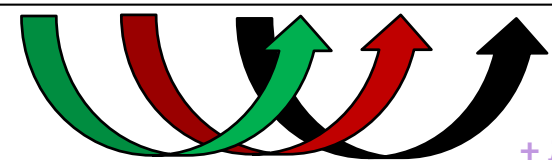
News Stories

Retractions
Corrections



Audience, Public Editors

Audience, Public Editors



+ Alerts. Filters.
+ Increased User Awareness.



LiT_{RL}

Victoria Rubin. CAIS/ACSI 2017

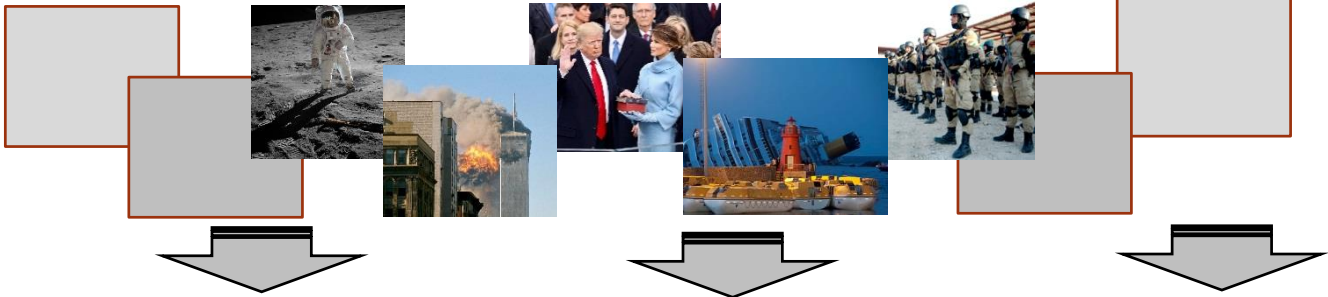
Mass Accountability

EXAMPLE

REPORTING A.S.A.P. MODEL

TECH. INTERVENTIONS

Events (World or Local)



Opportunities for text analysis to support human decision making

NEWS VERIFICATION SUITE

Sources (Direct Participants, Eye-Witnesses, Informants, etc.) + Experts + Previous Stories/Reports

Reporters:

Any falsehoods in eye witness accounts?
Previously reported? Similar events?

Editors:

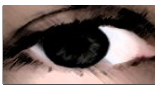
Is the style and rhetoric sensational?

Info curators:

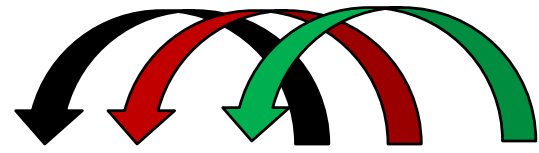
Any fraudulent information?
Suspicious stories?
Multiple perspectives?

News Readers

Content Creators ("Reporters a.s.a.p.")



Audience, Public Editors



Social Media Amplification



Mass Accountability

Deception Detection
Topicality Detection
Novelty Detection

Fraud Detection
Sentiment Analysis
Auto-Summarization

Rumor Debunking
Satire Labeling
Fact-Checking

+ Alerts. Filters.
+ Increased User Awareness.



LIMITATIONS

- **Limitations**
 - This vision requires extensive validation (while the problem is rampant and urgent)
 - Imperfect algorithms
 - Lacking accuracy. Built-in biases – solution is transparency & usability testing
 - NLP is primarily text-based solutions
 - Other complementary method for images, video and user behavior (looking into social networks & propagation)

FUTURE WORK

- Future work
 - interviews with journalists (to validate our insights & perceptions of news by readers)
 - development of individual modules (improving existing modules – Satire Detector, Deception Detection,; adding new ones - clickbait, native ads filter, fake news classification)
 - Streamlining the sequence of modules (fitting them into a reasonable set of tech interventions.)

CONCLUSIONS

The essence of my vision is in **integration of the tools into news reporting and news reading** for specific **NLP sub-tasks**, as exemplified by IFLA checklist, for example.

Let's pick one lock at a time!

THE NEWS VERIFICATION SUITE



To support news production + consumption

*To **unlock NLP potential** to support identification of Misinformation | Deception | "Fakes"*

HOW TO SPOT FAKE NEWS

- CONSIDER THE SOURCE**
Click away from the story to investigate the site, its mission and its contact info.
- READ BEYOND**
Headlines can be outrageous in an effort to get clicks. What's the whole story?
- CHECK THE AUTHOR**
Do a quick search on the author. Are they credible? Are they real?
- SUPPORTING SOURCES?**
Click on those links. Determine if the info given actually supports the story.
- CHECK THE DATE**
Reposting old news stories doesn't mean they're relevant to current events.
- IS IT A JOKE?**
If it is too outlandish, it might be satire. Research the site and author to be sure.
- CHECK YOUR BIASES**
Consider if your own beliefs could affect your judgement.
- ASK THE EXPERTS**
Ask a librarian, or consult a fact-checking site.

IFLA
International Federation of Library Associations and Institutions



REFERENCES

5. Acknowledgements

This research has been funded by the Government of Canada Social Sciences and Humanities Research Council (SSHRC) Insight Grant (#435-2015-0065) awarded to Dr. Rubin for the project entitled *Digital Deception Detection: Identifying Deliberate Misinformation in Online News*. This work could not have been possible without the dedicated team has been contributing towards a working system based on this vision. Special thanks for stimulating conversations to Paul Benedetti, an investigative journalist and a faculty member at Western, and my graduate research assistant team: Niall Conroy, Yimin Chen, Sarah Cornwell, and Toluwase Asubiaro.

6. References

- Chen, Y., Conroy, N. J., & Rubin, V. L. (2015a). *Misleading online content: Recognizing clickbait as "false news"*. Paper presented at the Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection.
- Chen, Y., Conroy, N. J., & Rubin, V. L. (2015b). *News in an online world: The need for an automatic crap detector*. Paper presented at the Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community.
- Cornwell, S. E. L., & Rubin, V. L. (2017). *Native advertising: Ads in disguise as editorials*. Paper presented at the 45th Annual Conference of the Canadian Association of Information Science, Toronto, Ontario, Canada.
- DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information systems research*, 3(1), 60-95. doi:10.1287/isre.3.1.60
- Donald trump's alarming alternative facts (2017, January 23, 2017). Editorial. *The Boston Globe*. Retrieved from <https://www.bostonglobe.com/opinion/editorials/2017/01/23/donald-trump-alarming-alternative-facts/sbmPy9m5wx4pMEzp4tdNnK/story.html>
- Fake News, Misinformation, and Propaganda. (2017). *Harvard Library Research Guide*. Retrieved from <http://guides.library.harvard.edu/fake>.
- Fox, C. J. (1983). Information and misinformation: An investigation of the notions of information, misinformation, informing, and misinforming. Westport, CT: Greenwood Press
- Galtung, J., & Ruge, M. (1973). Structuring and selecting news. *The manufacture of news: Social problems, deviance and the mass media*, 1(62), 62-72.
- Hancock, J. T. (2012). Digital deception: When, where and how people lie online. In A. N. Joinson, K. McKenna, T. Postmes, & U.-D. Reips (Eds.), *The oxford handbook of internet psychology* (pp. 508): Oxford University Press.
- Fighting fake news: How libraries can lead the way on media literacy (2016, December 27, 2016). *American Libraries Magazine*. Retrieved from <https://americanlibrariesmagazine.org/2016/12/27/fighting-fake-news/>
- How to spot fake news (2017, 18 April 2017). *IFLA Publications*. Retrieved from <https://www.ifla.org/publications/node/11174>
- How to spot fake news (2016, 18 November, 2016). *FactCheck.org Website*. Retrieved from <http://www.factcheck.org/2016/11/how-to-spot-fake-news/>
- Introducing Factmata: Artificial intelligence for automated fact-checking (2016, 6 December, 2016). *Factmata Website*. Retrieved from: <https://medium.com/factmata/introducing-factmata-artificial-intelligence-for-political-fact-checking-db8acdbf4cfl>
- Jarke, M., & Vassiliou, Y. (1997). *Data warehouse quality: A review of the dwq project*. Paper presented at the Conference on Information Quality, Massachusetts Institute of Technology, Cambridge MA USA.
- Knight, S.-a., & Burn, J. M. (2005). Developing a framework for assessing information quality on the world wide web. *Informing Science Journal*, 8(5), 159-172.
- Kovach, B., & Rosenstiel, T. (2010). *Blur: How to know what's true in the age of information overload* (Vol. 1st U.S.). New York: Bloomsbury.
- Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). Aimq: A methodology for information quality assessment. *Information & Management*, 40(2), 133-146.
- Rieh, S. Y. (2010). Credibility and cognitive authority of information. In M. J. Bates (Ed.), *Encyclopedia of library and information sciences* (Vol. 2, pp. 1337-1344): Taylor & Francis.
- Webinar: The Role of Information Scholars and Professionals in Responding to Fake News, Misinformation, and Propaganda. (2017, 31 March 2017). *ASIST Events*. Retrieved from: <https://www.asist.org/events/webinar-the-role-of-information-scholars-and-professionals-in-responding-to-fake-news-misinformation-and-propaganda/>
- Rubin, V. L. (2010). *On deception and deception detection: Content analysis of computer-mediated stated beliefs*. Paper presented at the 73rd ASIS&T Annual Meeting: Navigating Streams in an Information Ecosystem, Pittsburgh, Pennsylvania.
- Rubin, V. L. (2017). Deception Detection and Rumor Debunking for Social Media. In Sloan, L. & Quan-Haase, A. (Eds.) (2017) *The SAGE Handbook of Social Media Research Methods*, London: SAGE. <https://uk.sagepub.com/en-gb/eur/the-sage-handbook-of-social-media-research-methods/book245370>
- Rubin, V. L., Chen, Y., & Conroy, N. J. (2015). *Deception detection for news: Three types of fakes*. Paper presented at the Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community.
- Rubin, V. L., Conroy, N. J., & Chen, Y. (2015, January 5-8, 2015). *Towards news verification: Deception detection methods for news discourse*. Paper presented at the Hawaii International Conference on System Sciences (HICSS48) Symposium on Rapid Screening Technologies, Grand Hyatt, Kauai, Hawai'i.
- Rubin, V. L., Conroy, N. J., Chen, Y., & Cornwell, S. (2016, June 17, 2016). *Fake news or truth? Using satirical cues to detect potentially misleading news*. Paper presented at the The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA.
- Shoemaker, P. J., & Mayfield, E. K. (1987). Building a theory of news content: A synthesis of current approaches. *Journalism and Communication Monographs*, 103.
- Stvilia, B., Al-Faraj, A., & Yi, Y. J. (2009). Issues of cross-contextual information quality evaluation—the case of arabic, english, and korean wikipedias. *Library & Information Science Research*, 31(4), 232-239. doi:10.1016/j.lisr.2009.07.005
- Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11), 86-95. doi:10.1145/240455.240479
- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5-33.
- Zmud, R. W. (1978). An empirical investigation of the dimensionality of the concept of information. *Decision Sciences*, 9(2), 187-195. doi:10.1111/j.1540-5915.1978.tb01378.x

THANK YOU! SUGGESTIONS? QUESTIONS?



Contact Victoria Rubin:

vrubin@uwo.ca

<http://victoriarubin.fims.uwo.ca/>

Extended abstract available upon request.

FIMS

Faculty of Information & Media Studies

LIT **RL** Language & Information Technology Research Lab



<http://SatireDetector.fims.uwo.ca/>

Unsplash.com images