Nobody Said it Would be Easy: A Decade of R&D Projects in Information Access from Thomson over Reuters to Refinitiv

Jochen L. Leidner
leidner@acm.org
Refinitiv Labs / University of Sheffield
London / Sheffield

ABSTRACT

In this talk, I survey a small, non-random sample of research projects in information access carried out as part of the Thomson Reuters family of companies over the course of a 10+-year period.

I analyse into how these projects are similar and different when compared to academic research efforts and attempt a critical (and personal, so certainly subjective) assessment of what academia can do for industry, and what industry can do for research in terms of R&D efforts. I will conclude with some advice for academic-industry collaboration initiatives in several areas of vertical information services (legal, finance, pharma and regulatory/compliance) as well as news.

KEYWORDS

Information retrieval; information extraction; natural language processing; machine learning; innovation; academic collaboration; professional information services; corporate research & development

ACM Reference Format:
Jochen L. Leidner. 2019. Nobody Said it Would be Easy: A Decade of R&D Projects in Information Access from Thomson over Reuters to Refinitiv. In *42nd Int'l ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR'19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3331184.3331444

ABOUT THE AUTHOR

Dr. Jochen L. Leidner, M.A. M.Phil. Ph.D. F.R.G.S., is Director of Research, R&D at Refinitiv Labs, the innovation arm of Refinitiv Ltd., the company that used to be the Financial and Risk division of the Thomson Reuters Corporation.

After studying computational linguistics, English language and literature and computer science, he worked as a software developer at SAP before taking a Master's degree in Computer Speech, Text and Internet Technology at the University of Cambridge, England. He obtained a Ph.D. in information extraction at the University of Edinburgh, and after a postdoc as Royal Society of Edinburgh Enterprise Fellow in Electronic Markets pursuing research into the commercialization of mobile question answering, he joined the

Thomson Corporation as a Research Scientists. He helped set up and run innovation and research groups in Switzerland and the UK.

He is also Royal Academy of Engineering Visiting Professor of Data Analytics at the University of Sheffield and a Fellow of the Royal Geographical Society.

He was the recipient of the first ACM SIGIR Doctoral Consortium Award and twice winner of a Thomson Reuters "Inventor of the Year" award for the best patent application.

His research interests include information retrieval, knowledge management, natural language processing (in particular information extraction), geospatial information systems, applied machine learning and methodology.

ACKNOWLEDGMENTS

The author would like to thank his various project teams for the fascinating journey through various professional domains, for their technical brilliance, hard work and sense of humor.

REFERENCES

[1] Jochen Leidner and Frank Schilder. 2010. Hunting for the Black Swan: Risk Mining from Text. In *Proceedings of the ACL 2010 System Demonstrations.* Association for Computational Linguistics, Uppsala, Sweden, 54–59. https://www.aclweb.org/anthology/P10-4010

[2] Jochen L. Leidner and Vassilis Plachouras. 2017. Ethical by Design: Ethics Best Practices for Natural Language Processing. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing, EthNLP@EACL, Valencia, Spain, April 4, 2017*, Dirk Hovy, Shannon L. Spruit, Margaret Mitchell, Emily M. Bender, Michael Strube, and Hanna M. Wallach (Eds.). Association for Computational Linguistics, 30–40.

[3] Timothy Nugent and Jochen L. Leidner. 2016. Risk Mining: Company-Risk Identification from Unstructured Sources. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), 12-15 December, 2016.* IEEE, Barcelona, Spain, 1308–1311. http://ieeexplore.ieee.org/document/7836822/

[4] Fabio Petroni, Natraj Raman, Timothy Nugent, Armineh Nourbakhsh, Zarko Panic, Sameena Shah, and Jochen L. Leidner. 2018. An Extensible Event Extraction System

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SIGIR '19, July 21–25, 2019, Paris, France
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6172-9/19/07.
https://doi.org/10.1145/3331184.3331444

With Cross-Media Event Resolution. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, Yike Guo and Faisal Farooq (Eds.). ACM, New York, NY, USA, 626–635. https://doi.org/10.1145/3219819.3219827

[5] Daniel Pfeifer and Jochen L. Leidner. 2019. Topic Grouper: An Agglomerative Clustering Approach to Topic Modeling. In *Advances in Information Retrieval*, Leif Azzopardi, Benno Stein, Norbert Fuhr, Philipp Mayr, Claudia Hauff, and Djoerd Hiemstra (Eds.). Springer International Publishing, Cham, Switzerland, 590–603.

[6] Vassilis Plachouras, Jochen L. Leidner, and Andrew G. Garrow. 2016. Quantifying Self-Reported Adverse Drug Events on Twitter: Signal and Topic Analysis. In *Proceedings of the 7th 2016 International Conference on Social Media & Society, SMSociety 2016, London, UK, July 11 - 13, 2016*. ACM, 6:1–6:10. https://doi.org/10.1145/2930971.2930977

[7] Vassilis Plachouras, Charese Smiley, Hiroko Bretz, Ola Taylor, Jochen L. Leidner, Dezhao Song, and Frank Schilder. 2016. Interacting with Financial Data using Natural Language. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*, Raffaele Perego, Fabrizio Sebastiani, Javed A. Aslam, Ian Ruthven, and Justin Zobel (Eds.). ACM, 1121–1124. https://doi.org/10.1145/2911451.2911457

[8] Charese Smiley, Vassilis Plachouras, Frank Schilder, Hiroko Bretz, Jochen L. Leidner, and Dezhao Song. 2016. When to Plummet and When to Soar: Corpus Based Verb Selection for Natural Language Generation. In *INLG 2016 - Proceedings of the Ninth International Natural Language Generation Conference, September 5-8, 2016, Edinburgh, UK*, Amy Isard, Verena Rieser, and Dimitra Gkatzia (Eds.). The Association for Computer Linguistics, New York, NY, USA, 36–39.