An Introduction to Neural Information Retrieval

Other titles in Foundations and Trends® in Information Retrieval

Efficient Query Processing for Scalable Web Search
Nicola Tonellotto, Craig Macdonald and Iadh Ounis
ISBN: 978-1-68083-542-7

Geographic Information Retrieval: Progress and Challenges in Spatial Search of Text
Ross S. Purves, Paul Clough, Christopher B. Jones, Mark H. Hall and Vanessa Murdock
ISBN: 978-1-68083-412-3

Web Forum Retrieval and Text Analytics: A Survey
Doris Hoogeveen, Li Wang, Timothy Baldwin and Karin M. Verspoor
ISBN: 978-1-68083-350-8

An Introduction to Neural Information Retrieval

Bhaskar Mitra
Microsoft, University College London
Montreal, Canada
bmitra@microsoft.com

Nick Craswell
Microsoft
Bellevue, USA
nickcr@microsoft.com

the essence of knowledge
Boston — Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

B. Mitra and N. Craswell. *An Introduction to Neural Information Retrieval*. Foundations and Trends® in Information Retrieval, vol. 13, no. 1, pp. 1–126, 2018.

ISBN: 978-1-68083-533-5

Foundations and Trends® in Information Retrieval

Volume 13, Issue 1, 2019

Editorial Scope

Topics

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2019, Volume 13, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Contents

An Introduction to Neural Information Retrieval

Bhaskar Mitra[1] and Nick Craswell[2]

[1] *Microsoft, University College London; bmitra@microsoft.com*
[2] *Microsoft; nickcr@microsoft.com*

ABSTRACT

Neural ranking models for information retrieval (IR) use shallow or deep neural networks to rank search results in response to a query. Traditional learning to rank models employ supervised machine learning (ML) techniques—including neural networks—over hand-crafted IR features. By contrast, more recently proposed neural models learn representations of language from raw text that can bridge the gap between query and document vocabulary. Unlike classical learning to rank models and non-neural approaches to IR, these new ML techniques are data-hungry, requiring large scale training data before they can be deployed. This tutorial introduces basic concepts and intuitions behind neural IR models, and places them in the context of classical non-neural approaches to IR. We begin by introducing fundamental concepts of retrieval and different neural and non-neural approaches to unsupervised learning of vector representations of text. We then review IR methods that employ these pre-trained neural vector representations without learning the IR task end-to-end. We introduce the Learning to Rank (LTR) framework next, discussing standard loss functions for ranking. We follow that with an overview of deep neural networks (DNNs), including standard architectures and implementations. Finally, we review supervised neural learning to rank models, including recent DNN architectures trained end-to-end for ranking tasks. We conclude with a discussion on potential future directions for neural IR.

Bhaskar Mitra and Nick Craswell (2018), “An Introduction to Neural Information Retrieval”, Foundations and Trends® in Information Retrieval: Vol. 13, No. 1, pp 1–126. DOI: 10.1561/1500000061.

1

Introduction

Since the turn of the decade, there have been dramatic improvements in performance in computer vision, speech recognition, and machine translation tasks, witnessed in research and in real-world applications (LeCun *et al.*, 2015). These breakthroughs were largely fuelled by recent advances in neural network models, usually with multiple hidden layers, known as deep architectures (Krizhevsky *et al.*, 2012; LeCun *et al.*, 2015; Hinton *et al.*, 2012; Bahdanau *et al.*, 2014; Deng, Yu, *et al.*, 2014) combined with the availability of large datasets (Wissner-Gross, 2016) and cheap compute power for model training. Exciting novel applications, such as conversational agents (Vinyals and Le, 2015; Sordoni *et al.*, 2015b), have also emerged, as well as game-playing agents with human-level performance (Silver *et al.*, 2016; Mnih *et al.*, 2015). Work has now begun in the information retrieval (IR) community to apply these neural methods, leading to the possibility of advancing the state of the art or even achieving breakthrough performance as in these other fields.

Retrieval of information can take many forms (White, 2016). Users can express their information need in the form of a text query—by typing on a keyboard, by selecting a query suggestion, or by voice recognition—or the query can be in the form of an image, or in some cases the need can be implicit. Retrieval can involve ranking existing pieces of content, such as documents or short-text answers, or composing new responses incorporating retrieved information. Both the information need and the retrieved results may use the same modality (*e.g.*, retrieving text documents in response to keyword queries), or be different (*e.g.*, image search using text queries). If the query is ambiguous, retrieval system may consider user history, physical location, temporal changes in information, or other context when ranking results. IR systems may also help users formulate their intent (*e.g.*, via query auto-completion or query suggestion) and can extract succinct summaries of results that take the user's query into account.

We note that many natural language processing tasks exist that are not IR. Machine translation of text from one human language to another is not an IR task, because translating language and searching a corpus to satisfy a user's information need are different. However, translation could be used in an IR system, to enable cross-language retrieval on a multilingual corpus (Oard and Diekema, 1998). Named entity linking, where text is disambiguated through linking to a knowledgebase, is not an IR task in itself. However, an IR system could use entity linking to enhance its performance on IR tasks. In general, many natural language processing tasks do not involve information access and retrieval, so are not IR tasks, but some can still be useful as part of a larger IR system.

> Neural IR is the application of shallow or deep neural networks to IR tasks. Other natural language processing capabilities such as machine translation and named entity linking are *not* neural IR but could be used in an IR system.

Neural IR refers to the application of shallow or deep neural networks to retrieval tasks. Neural models have been employed in many

IR scenarios—including ad-hoc retrieval, recommender systems, multimedia search, and even conversational systems that generate answers in response to natural language questions. This tutorial serves as an introduction to neural methods for ranking documents in response to a query, an important IR task. We scope our discussions to a single task to allow for more thorough treatment of the fundamentals as opposed to providing a shallow survey of neural approaches to all IR tasks.

A search query may typically contain a few terms, while the document length, depending on the scenario, may range from a few terms to hundreds of sentences or more. Neural models for IR use vector representations of text, and usually contain a large number of parameters that need to be tuned. ML models with large set of parameters typically benefit from large quantity of training data (Brill, 2003; Taylor *et al.*, 2006; Rajaraman, 2008; Halevy *et al.*, 2009; Sun *et al.*, 2017). Unlike traditional *learning to rank* (LTR) approaches (Liu, 2009) that train ML models over a set of hand-crafted features, recent neural models for IR typically accept the raw text of a query and document as input. Learning suitable representations of text also demands large-scale datasets for training (Mitra *et al.*, 2017a). Therefore, unlike classical IR models, these neural approaches tend to be data hungry, with performance that improves with more training data.

Text representations can be learnt in an unsupervised or supervised fashion. The supervised approach uses IR data such as labelled query-document pairs, to learn a representation that is optimized end-to-end for the task at hand. If sufficient IR labels are not available, the unsupervised approach learns a representation using just the queries and/or documents. In the latter case, different unsupervised learning setups may lead to vector representations that capture different notions of text similarity. When applying such representations, the choice of unsupervised learning setup should be carefully considered, to yield a notion of text similarity that is suitable for the target task. Traditional IR models such as *Latent Semantic Analysis* (LSA) (Deerwester *et al.*, 1990) learn dense vector representations of terms and documents. Neural representation learning models share commonalities with these traditional approaches. Much of our understanding of these traditional

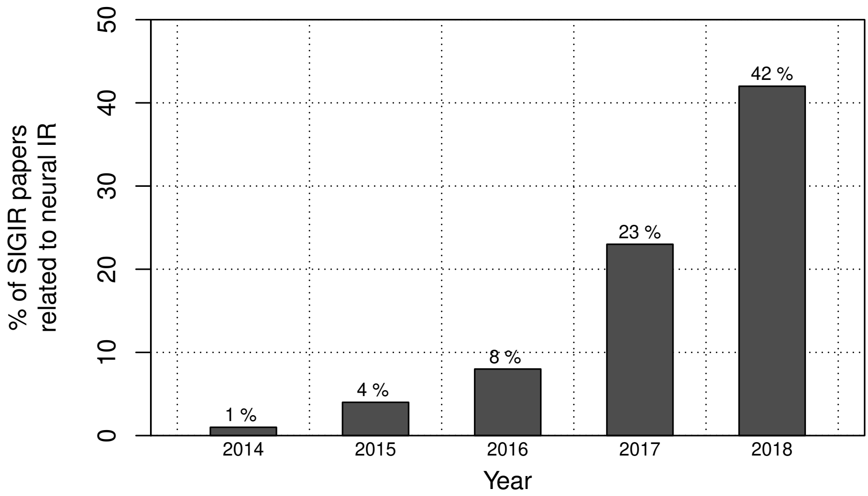

Figure 1.1: The percentage of neural IR papers at the ACM SIGIR conference—as determined by a manual inspection of the papers—shows a clear trend in the growing popularity of the field.

approaches from decades of research can be extended to these modern representation learning models.

In other fields, the design of neural network models has been informed by characteristics of the application and data. For example, the datasets and successful architectures are quite different in visual object recognition, speech recognition, and game playing agents. While IR shares some common attributes with the field of natural language processing, it also comes with its own set of unique challenges. IR systems must deal with short queries that may contain previously unseen vocabulary, to match against documents that vary in length, to find relevant documents that may also contain large sections of irrelevant text. IR systems should learn patterns in query and document text that indicate relevance, even if query and document use different vocabulary, and even if the patterns are task-specific or context-specific.

The goal of this tutorial is to introduce the fundamentals of neural IR, in context of traditional IR research, with visual examples to illustrate key concepts and a consistent mathematical notation for describing key models. Section 2 presents a survey of IR tasks, challenges, metrics and non-neural models—as well as a brief overview of different neural approaches to IR. Section 3 introduces neural and non-neural methods for learning term embeddings, without the use of supervision from IR labels, and with a focus on the notions of similarity. Section 4 surveys some specific approaches for incorporating such unsupervised embeddings in IR. Section 5 introduces supervised learning to rank models. Section 6 introduces the fundamentals of deep models—including standard architectures and toolkits—before §7 surveys some specific approaches for incorporating deep neural networks (DNNs) in IR. Section 8 is our discussion, including future work, and conclusion.

Motivation for this tutorial Neural IR is an emerging field. Research publication in the area has been increasing (Figure 1.1), along with relevant workshops (Craswell *et al.*, 2016a; Craswell *et al.*, 2016b; Craswell *et al.*, 2017; Craswell *et al.*, 2018), tutorials (Li and Lu, n.d.; Mitra and Craswell, 2017; Kenter *et al.*, 2017; Kenter *et al.*, 2018a; Kenter *et al.*, 2018b), and plenary talks (Manning, 2016; Craswell, 2017). Because this growth in interest is fairly recent, some researchers with IR expertise may be unfamiliar with neural models, and other researchers who have already worked with neural models may be unfamiliar with IR. The purpose of this tutorial is to bridge the gap, by describing the relevant IR concepts and neural methods in the current literature.

A thorough review of the fundamentals of IR or neural networks, however, are beyond the scope of this tutorial. We refer interested readers to the following books for a general overview of traditional IR.

1. *Modern information retrieval*, by Baeza-Yates and Ribeiro-Neto (1999).

2. *Introduction to information retrieval*, by Manning *et al.* (2008).

3. *Search engines: Information retrieval in practice*, by Croft *et al.* (2010).

Similarly, we recommend the following as companion reading materials for machine learning and neural network fundamentals.

1. *The elements of statistical learning*, by Hastie *et al.* (2001).

2. *Pattern recognition and machine learning*, by Bishop (2006).
3. *Deep learning*, by Goodfellow *et al.* (2016).

References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, O. Viegas Fernanda and-Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. 2016. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems". *arXiv preprint arXiv:1603.04467*.

Abdul-Jaleel, N., J. Allan, W. B. Croft, F. Diaz, L. Larkey, X. Li, M. D. Smucker, and C. Wade. 2004. "UMass at TREC 2004: Novelty and HARD".

Agichtein, E., D. Carmel, D. Pelleg, Y. Pinter, and D. Harman. 2015. "Overview of the TREC 2015 LiveQA Track." In: *TREC*.

Ai, Q., L. Yang, J. Guo, and W. B. Croft. 2016a. "Analysis of the paragraph vector model for information retrieval". In: *Proc. ICTIR*. ACM. 133–142.

Ai, Q., L. Yang, J. Guo, and W. B. Croft. 2016b. "Improving language estimation with the paragraph vector model for ad-hoc retrieval". In: *Proc. SIGIR*. ACM. 869–872.

Arora, S., Y. Li, Y. Liang, T. Ma, and A. Risteski. 2015. "Rand-walk: A latent variable model approach to word embeddings". *arXiv preprint arXiv:1502.03520*.

Baeza-Yates, R. and B. Ribeiro-Neto. 1999. *Modern information retrieval*. Vol. 463. ACM press New York.

Bahdanau, D., K. Cho, and Y. Bengio. 2014. "Neural machine translation by jointly learning to align and translate". *arXiv preprint arXiv:1409.0473*.

Bahuleyan, H., L. Mou, O. Vechtomova, and P. Poupart. 2017. "Variational Attention for Sequence-to-Sequence Models". *arXiv preprint arXiv:1712.08207*.

Bailey, P., N. Craswell, and D. Hawking. 2003. "Engineering a multi-purpose test collection for web retrieval experiments". *Information Processing & Management*. 39(6): 853–871.

Baldi, P. and Y. Chauvin. 1993. "Neural networks for fingerprint recognition". *Neural Computation*. 5(3): 402–418.

Baroni, M., G. Dinu, and G. Kruszewski. 2014. "Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors". In: *Proc. ACL*. Vol. 1. 238–247.

Baroni, M. and A. Lenci. 2010. "Distributional memory: A general framework for corpus-based semantics". *Computational Linguistics*. 36(4): 673–721.

Barthes, R. 1977. *Elements of semiology*. Macmillan.

Benetka, J. R., K. Balog, and K. Nørvåg. 2017. "Anticipating Information Needs Based on Check-in Activity". In: *WSDM*. ACM. 41–50.

Bengio, Y. 2009. "Learning deep architectures for AI". *Foundations and trends® in Machine Learning*. 2(1): 1–127.

Bengio, Y., R. Ducharme, P. Vincent, and C. Jauvin. 2003. "A neural probabilistic language model". *Journal of machine learning research*. 3(Feb): 1137–1155.

Bengio, Y., P. Lamblin, D. Popovici, H. Larochelle, *et al.* 2007. "Greedy layer-wise training of deep networks". *Proc. NIPS*. 19: 153.

Bengio, Y. and J.-S. Senécal. 2008. "Adaptive importance sampling to accelerate training of a neural probabilistic language model". *IEEE Transactions on Neural Networks*. 19(4): 713–722.

Bengio, Y., J.-S. Senécal, *et al.* 2003. "Quick Training of Probabilistic Neural Nets by Importance Sampling." In: *AISTATS*.

Berger, A. and J. Lafferty. 1999. "Information retrieval as statistical translation". In: *Proc. SIGIR*. ACM. 222–229.

Bhatia, K., P. Jain, and M. Varma. 2016. "The extreme classification repository: multi-label datasets & code". Accessed June 7, 2017. URL: http://manikvarma.org/downloads/XC/XMLRepository.html.

Bishop, C. M. 2006. *Pattern recognition and machine learning*. Springer-Verlag New York.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent dirichlet allocation". *the Journal of machine Learning research*. 3: 993–1022.

Bordes, A., N. Usunier, S. Chopra, and J. Weston. 2015. "Large-scale simple question answering with memory networks". *arXiv preprint arXiv:1506.02075*.

Borisov, A., I. Markov, M. de Rijke, and P. Serdyukov. 2016a. "A context-aware time model for web search". In: *Proc. SIGIR*. ACM. 205–214.

Borisov, A., I. Markov, M. de Rijke, and P. Serdyukov. 2016b. "A neural click model for web search". In: *Proc. WWW*. 531–541.

Bowman, S. R., J. Gauthier, A. Rastogi, R. Gupta, C. D. Manning, and C. Potts. 2016. "A fast unified model for parsing and sentence understanding". *arXiv preprint arXiv:1603.06021*.

Brébisson, A. de and P. Vincent. 2015. "An exploration of softmax alternatives belonging to the spherical loss family". *arXiv preprint arXiv:1511.05042*.

Brébisson, A. de and P. Vincent. 2016. "The Z-loss: a shift and scale invariant classification loss belonging to the Spherical Family". *arXiv preprint arXiv:1604.08859*.

Brill, E. 2003. "Processing natural language without natural language processing". In: *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer. 360–369.

Bromley, J., J. W. Bentz, L. Bottou, I. Guyon, Y. LeCun, C. Moore, E. Säckinger, and R. Shah. 1993. "Signature Verification Using A "Siamese" Time Delay Neural Network". *IJPRAI*. 7(4): 669–688.

Brown, P. F., J. Cocke, S. A. D. Pietra, V. J. D. Pietra, F. Jelinek, J. D. Lafferty, R. L. Mercer, and P. S. Roossin. 1990. "A statistical approach to machine translation". *Computational linguistics*. 16(2): 79–85.

Brown, P. F., P. V. Desouza, R. L. Mercer, V. J. D. Pietra, and J. C. Lai. 1992. "Class-based n-gram models of natural language". *Computational linguistics*. 18(4): 467–479.

Brown, P. F., V. J. D. Pietra, S. A. D. Pietra, and R. L. Mercer. 1993. "The mathematics of statistical machine translation: Parameter estimation". *Computational linguistics*. 19(2): 263–311.

Buck, C., J. Bulian, M. Ciaramita, A. Gesmundo, N. Houlsby, W. Gajewski, and W. Wang. 2017. "Ask the right questions: Active question reformulation with reinforcement learning". *arXiv preprint arXiv:1705.07830*.

Bullinaria, J. A. and J. P. Levy. 2007. "Extracting semantic representations from word co-occurrence statistics: A computational study". *Behavior research methods*. 39(3): 510–526.

Bullinaria, J. A. and J. P. Levy. 2012. "Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and SVD". *Behavior research methods*. 44(3): 890–907.

Burges, C., T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. 2005. "Learning to rank using gradient descent". In: *Proc. ICML*. ACM. 89–96.

Burges, C. J. 2010. "From ranknet to lambdarank to lambdamart: An overview". *Learning*. 11(23-581): 81.

Burges, C. J., R. Ragno, and Q. V. Le. 2006. "Learning to rank with nonsmooth cost functions". In: *NIPS*. Vol. 6. 193–200.

Callan, J., M. Hoy, C. Yoo, and L. Zhao. 2009. "Clueweb09 data set".

Cao, Z., T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li. 2007. "Learning to rank: from pairwise approach to listwise approach". In: *Proc. ICML*. ACM. 129–136.

Chaidaroon, S. and Y. Fang. 2017. "Variational Deep Semantic Hashing for Text Documents". In: *Proc. SIGIR*.

Chandler, D. 1994. "Semiotics for beginners".

Chen, T., M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang. 2015a. "Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems". *arXiv preprint arXiv:1512.01274*.

Chen, W., T.-Y. Liu, Y. Lan, Z.-M. Ma, and H. Li. 2009. "Ranking measures and loss functions in learning to rank". In: *Advances in Neural Information Processing Systems*. 315–323.

Chen, W., D. Grangier, and M. Auli. 2015b. "Strategies for training large vocabulary neural language models". *arXiv preprint arXiv:1512.04906*.

Cho, K., B. Van Merriënboer, D. Bahdanau, and Y. Bengio. 2014. "On the properties of neural machine translation: Encoder-decoder approaches". *arXiv preprint arXiv:1409.1259*.

Chopra, S., R. Hadsell, and Y. LeCun. 2005. "Learning a similarity metric discriminatively, with application to face verification". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE. 539–546.

Chorowski, J. K., D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio. 2015. "Attention-based models for speech recognition". In: *Proc. NIPS*. 577–585.

Chung, J., C. Gulcehre, K. Cho, and Y. Bengio. 2014. "Empirical evaluation of gated recurrent neural networks on sequence modeling". *arXiv preprint arXiv:1412.3555*.

Cisse, M., M. Varma, and S. Bengio. 2016. "Extreme Classification 2016: The NIPS Workshop on Multi-class and Multi-label Learning in Extremely Large Label Spaces". Accessed June 7, 2017. URL: http://manikvarma.org/events/XC16/.

Clinchant, S. and F. Perronnin. 2013. "Aggregating continuous word embeddings for information retrieval". In: *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*. 100–109.

Cohen, D., Q. Ai, and W. B. Croft. 2016. "Adaptability of neural networks on varying granularity IR tasks". *arXiv preprint arXiv:1606.07565*.

Cohen, D. and W. B. Croft. 2016. "End to End Long Short Term Memory Networks for Non-Factoid Question Answering". In: *Proc. ICTIR*. ACM. 143–146.

Cohen, D., B. Mitra, K. Hofmann, and B. Croft. 2018. "Cross Domain Regularization for Neural Ranking Models using Adversarial Learning". In: *Proc. SIGIR*. ACM.

Collobert, R., K. Kavukcuoglu, and C. Farabet. 2011a. "Torch7: A matlab-like environment for machine learning". In: *BigLearn, NIPS Workshop*. No. EPFL-CONF-192376.

Collobert, R., J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. 2011b. "Natural language processing (almost) from scratch". *The Journal of Machine Learning Research*. 12: 2493–2537.

Cossock, D. and T. Zhang. 2006. "Subset ranking using regression". In: *COLT*. Vol. 6. Springer. 605–619.

Craswell, N. 2009. "Mean reciprocal rank". In: *Encyclopedia of Database Systems*. Springer. 1703–1703.

Craswell, N. 2017. "Neural Models for Full Text Search". In: *Proc. WSDM*. ACM. 251–251.

Craswell, N., W. B. Croft, J. Guo, B. Mitra, and M. de Rijke. 2016a. "Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval".

Craswell, N., W. B. Croft, J. Guo, B. Mitra, and M. de Rijke. 2016b. "Report on the SIGIR 2016 Workshop on Neural Information Retrieval (Neu-IR)". *ACM Sigir forum*. 50(2): 96–103.

Craswell, N., W. B. Croft, M. de Rijke, J. Guo, and B. Mitra. 2017. "Neu-IR'17: Neural Information Retrieval". In: *Proc. SIGIR*. ACM.

Craswell, N., W. B. Croft, M. de Rijke, J. Guo, and B. Mitra. 2018. "Report on the Second SIGIR Workshop on Neural Information Retrieval (Neu-IR'17)". In: *ACM SIGIR Forum*. Vol. 51. No. 3. ACM. 152–158.

Craswell, N., D. Hawking, R. Wilkinson, and M. Wu. 2003. "Overview of the TREC 2002 Web track". In: *TREC*. Vol. 3. 12th.

Craswell, N., R. Jones, G. Dupret, and E. Viegas. 2009. *Proceedings of the 2009 workshop on Web Search Click Data*. ACM.

Croft, W. B., D. Metzler, and T. Strohman. 2010. *Search engines: Information retrieval in practice*. Vol. 283. Addison-Wesley Reading.

De Saussure, F. 1916. "Cours de linguistique générale, publié par Ch". *Bally et A. Sechehaye avec la collaboration de A. Riedlinger. Paris: Payot*.

Deerwester, S. C., S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. 1990. "Indexing by latent semantic analysis". *JASIS*. 41(6): 391–407.

Dehghani, M., A. Severyn, S. Rothe, and J. Kamps. 2017a. "Learning to Learn from Weak Supervision by Full Supervision". *arXiv preprint arXiv:1711.11383*.

Dehghani, M., H. Zamani, A. Severyn, J. Kamps, and W. B. Croft. 2017b. "Neural ranking models with weak supervision". In: *Proc. SIGIR*. ACM. 65–74.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the royal statistical society. Series B (methodological)*: 1–38.

Deng, L., D. Yu, *et al.* 2014. "Deep learning: methods and applications". *Foundations and Trends® in Signal Processing*. 7(3–4): 197–387.

Denil, M., A. Demiraj, N. Kalchbrenner, P. Blunsom, and N. de Freitas. 2014. "Modelling, visualising and summarising documents with a single convolutional neural network". *arXiv preprint arXiv:1406.3830*.

Diaz, F., B. Mitra, and N. Craswell. 2016. "Query Expansion with Locally-Trained Word Embeddings". In: *Proc. ACL*.

Diaz, F., R. White, G. Buscher, and D. Liebling. 2013. "Robust models of mouse movement on dynamic web search results pages". In: *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM. 1451–1460.

Donmez, P., K. M. Svore, and C. J. Burges. 2009. "On the local optimality of LambdaRank". In: *Proc. SIGIR*. ACM. 460–467.

Duchi, J., E. Hazan, and Y. Singer. 2011. "Adaptive subgradient methods for online learning and stochastic optimization". *Journal of Machine Learning Research*. 12(Jul): 2121–2159.

Dyer, C. 2014. "Notes on Noise Contrastive Estimation and Negative Sampling". *arXiv preprint arXiv:1410.8251*.

Elman, J. L. 1990. "Finding structure in time". *Cognitive science*. 14(2): 179–211.

Ferrucci, D., E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, *et al.* 2010. "Building Watson: An overview of the DeepQA project". *AI magazine*. 31(3): 59–79.

Firth, J. R. 1957. "A synopsis of linguistic theory, 1930–1955". In: *Studies in Linguistic Analysis*. Blackwell. 1–32.

Freund, Y., R. Iyer, R. E. Schapire, and Y. Singer. 2003. "An efficient boosting algorithm for combining preferences". *Journal of machine learning research*. 4(Nov): 933–969.

Fuhr, N. 1989. "Optimum polynomial retrieval functions based on the probability ranking principle". *ACM Transactions on Information Systems (TOIS)*. 7(3): 183–204.

Galley, M., C. Brockett, A. Sordoni, Y. Ji, M. Auli, C. Quirk, M. Mitchell, J. Gao, and B. Dolan. 2015. "deltaBLEU: A discriminative metric for generation tasks with intrinsically diverse targets". *arXiv preprint arXiv:1506.06863*.

Ganguly, D., D. Roy, M. Mitra, and G. J. Jones. 2015. "Word Embedding based Generalized Language Model for Information Retrieval". In: *Proc. SIGIR*. ACM. 795–798.

Gao, J., P. Pantel, M. Gamon, X. He, L. Deng, and Y. Shen. 2014. "Modeling interestingness with deep neural networks". In: *Proc. EMNLP*.

Ghazvininejad, M., C. Brockett, M.-W. Chang, B. Dolan, J. Gao, W.-t. Yih, and M. Galley. 2017. "A Knowledge-Grounded Neural Conversation Model". *arXiv preprint arXiv:1702.01932*.

Goldberg, Y. and O. Levy. 2014. "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method". *arXiv preprint arXiv:1402.3722*.

Goller, C. and A. Kuchler. 1996. "Learning task-dependent distributed representations by backpropagation through structure". In: *Neural Networks, 1996., IEEE International Conference on*. Vol. 1. IEEE. 347–352.

Golub, G. H. and C. Reinsch. 1970. "Singular value decomposition and least squares solutions". *Numerische mathematik*. 14(5): 403–420.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. *Deep learning*. MIT Press.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. "Generative adversarial nets". In: *Proc. NIPS*. 2672–2680.

Goodman, J. 2001. "Classes for fast maximum entropy training". In: *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP'01). 2001 IEEE International Conference on*. Vol. 1. IEEE. 561–564.

Granka, L. A., T. Joachims, and G. Gay. 2004. "Eye-tracking analysis of user behavior in WWW search". In: *Proc. SIGIR*. ACM. 478–479.

Grave, E., A. Joulin, M. Cissé, D. Grangier, and H. Jégou. 2016. "Efficient softmax approximation for GPUs". *arXiv preprint arXiv:1609.04309*.

Graves, A. 2013. "Generating sequences with recurrent neural networks". *arXiv preprint arXiv:1308.0850*.

Graves, A., M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber. 2009. "A novel connectionist system for unconstrained handwriting recognition". *IEEE transactions on pattern analysis and machine intelligence*. 31(5): 855–868.

Grbovic, M., N. Djuric, V. Radosavljevic, and N. Bhamidipati. 2015a. "Search Retargeting using Directed Query Embeddings". In: *Proc. WWW*. International World Wide Web Conferences Steering Committee. 37–38.

Grbovic, M., N. Djuric, V. Radosavljevic, F. Silvestri, and N. Bhamidipati. 2015b. "Context-and Content-aware Embeddings for Query Rewriting in Sponsored Search". In: *Proc. SIGIR*. ACM. 383–392.

Gregor, K., I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra. 2015. "DRAW: A recurrent neural network for image generation". *arXiv preprint arXiv:1502.04623*.

Guan, Z. and E. Cutrell. 2007. "An eye tracking study of the effect of target rank on web search". In: *Proc. SIGCHI*. ACM. 417–420.

Guo, J., Y. Fan, Q. Ai, and W. B. Croft. 2016a. "A Deep Relevance Matching Model for Ad-hoc Retrieval". In: *Proc. CIKM*. ACM. 55–64.

Guo, J., Y. Fan, Q. Ai, and W. B. Croft. 2016b. "Semantic Matching by Non-Linear Word Transportation for Information Retrieval". In: *Proc. CIKM*. ACM. 701–710.

Gutmann, M. and A. Hyvärinen. 2010. "Noise-contrastive estimation: A new estimation principle for unnormalized statistical models." In: *AISTATS*. Vol. 1. No. 2. 6.

Hadsell, R., S. Chopra, and Y. LeCun. 2006. "Dimensionality reduction by learning an invariant mapping". In: *Computer vision and pattern recognition, 2006 IEEE computer society conference on*. Vol. 2. IEEE. 1735–1742.

Hahnloser, R. H., R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung. 2000. "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit". *Nature*. 405(6789): 947–951.

Halevy, A., P. Norvig, and F. Pereira. 2009. "The unreasonable effectiveness of data". *IEEE Intelligent Systems*. 24(2): 8–12.

Harris, R. 2001. *Saussure and his Interpreters*. Edinburgh University Press.

Harris, Z. S. 1954. "Distributional structure". *Word*. 10(2-3): 146–162.

Hastie, T., R. Tibshirani, and J. Friedman. 2001. "The elements of statistical learning. 2001".

He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep residual learning for image recognition". In: *Proc. CVPR*. 770–778.

Hecht-Nielsen, R. *et al.* 1988. "Theory of the backpropagation neural network." *Neural Networks*. 1(Supplement-1): 445–448.

Herbrich, R., T. Graepel, and K. Obermayer. 2000. "Large margin rank boundaries for ordinal regression".

Hermann, K. M., T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom. 2015. "Teaching machines to read and comprehend". In: *Proc. NIPS*. 1693–1701.

Hiemstra, D. 2001. *Using language models for information retrieval.* Taaluitgeverij Neslia Paniculata.

Hill, F., A. Bordes, S. Chopra, and J. Weston. 2015. "The Goldilocks Principle: Reading Children's Books with Explicit Memory Representations". *arXiv preprint arXiv:1511.02301*.

Hinton, G. E. 1984. "Distributed representations".

Hinton, G., L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, *et al.* 2012. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups". *Signal Processing Magazine, IEEE*. 29(6): 82–97.

Hochreiter, S. and J. Schmidhuber. 1997. "Long short-term memory". *Neural computation*. 9(8): 1735–1780.

Hofmann, K., B. Mitra, F. Radlinski, and M. Shokouhi. 2014. "An Eye-tracking Study of User Interactions with Query Auto Completion". In: *Proc. CIKM*. ACM. 549–558.

Hofmann, T. 1999. "Probabilistic latent semantic indexing". In: *Proc. SIGIR*. ACM. 50–57.

Hornik, K., M. Stinchcombe, and H. White. 1989. "Multilayer feedforward networks are universal approximators". *Neural networks*. 2(5): 359–366.

Hu, B., Z. Lu, H. Li, and Q. Chen. 2014. "Convolutional neural network architectures for matching natural language sentences". In: *Proc. NIPS*. 2042–2050.

Huang, G., C. Guo, M. J. Kusner, Y. Sun, F. Sha, and K. Q. Weinberger. 2016. "Supervised Word Mover's Distance". In: *Proc. NIPS*. 4862–4870.

Huang, P.-S., X. He, J. Gao, L. Deng, A. Acero, and L. Heck. 2013. "Learning deep structured semantic models for web search using clickthrough data". In: *Proc. CIKM*. ACM. 2333–2338.

Hui, K., A. Yates, K. Berberich, and G. de Melo. 2017. "PACRR: A Position-Aware Neural IR Model for Relevance Matching". In: *Proc. EMNLP*. 1049–1058.

Hui, K., A. Yates, K. Berberich, and G. de Melo. 2018. "Co-PACRR: A Context-Aware Neural IR Model for Ad-hoc Retrieval". In: *Proceedings of the 11th ACM International Conference on Web Search and Data Mining. WSDM*. Vol. 18. 2.

Jain, H., Y. Prabhu, and M. Varma. 2016. "Extreme Multi-label Loss Functions for Recommendation, Tagging, Ranking & Other Missing Label Applications". In: *Proc. SIGKDD*. ACM. 935–944.

Jarrett, K., K. Kavukcuoglu, Y. LeCun, *et al.* 2009. "What is the best multi-stage architecture for object recognition?" In: *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE. 2146–2153.

Järvelin, K. and J. Kekäläinen. 2002. "Cumulated gain-based evaluation of IR techniques". *ACM Transactions on Information Systems (TOIS)*. 20(4): 422–446.

Jean, S., K. Cho, R. Memisevic, and Y. Bengio. 2014. "On Using Very Large Target Vocabulary for Neural Machine Translation". *arXiv preprint arXiv:1412.2007*.

Jelinek, F. and R. Mercer. 1980. "Interpolated estimation of Markov source parameters from sparse data". In: *Proc. Workshop on Pattern Recognition in Practice, 1980*.

Ji, S., S. Vishwanathan, N. Satish, M. J. Anderson, and P. Dubey. 2015. "Blackout: Speeding up recurrent neural network language models with very large vocabularies". *arXiv preprint arXiv:1511.06909*.

Jia, Y., E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. 2014. "Caffe: Convolutional architecture for fast feature embedding". In: *Proceedings of the 22nd ACM international conference on Multimedia*. ACM. 675–678.

Joachims, T., L. Granka, B. Pan, H. Hembrooke, and G. Gay. 2005. "Accurately interpreting clickthrough data as implicit feedback". In: *Proc. SIGIR*. Acm. 154–161.

Joachims, T., L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay. 2007. "Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search". *ACM Transactions on Information Systems (TOIS)*. 25(2): 7.

Joachims, T., A. Swaminathan, and T. Schnabel. 2017. "Unbiased learning-to-rank with biased feedback". In: *Proc. WSDM*. ACM. 781–789.

Jozefowicz, R., O. Vinyals, M. Schuster, N. Shazeer, and Y. Wu. 2016. "Exploring the limits of language modeling". *arXiv preprint arXiv:1602.02410*.

Kalchbrenner, N., E. Grefenstette, and P. Blunsom. 2014. "A convolutional neural network for modelling sentences". *arXiv preprint arXiv:1404.2188*.

Kenter, T., A. Borisov, and M. de Rijke. 2016. "Siamese cbow: Optimizing word embeddings for sentence representations". *arXiv preprint arXiv:1606.04640*.

Kenter, T., A. Borisov, C. Van Gysel, M. Dehghani, M. de Rijke, and B. Mitra. 2017. "Neural Networks for Information Retrieval (NN4IR)". In: *Proc. SIGIR*. ACM.

Kenter, T., A. Borisov, C. Van Gysel, M. Dehghani, M. de Rijke, and B. Mitra. 2018a. "Neural networks for information retrieval". In: *Proc. WSDM*. ACM. 779–780.

Kenter, T., A. Borisov, C. Van Gysel, M. Dehghani, M. de Rijke, and B. Mitra. 2018b. "Neural networks for information retrieval". In: *Proc. ECIR*.

Kenter, T. and M. de Rijke. 2015. "Short Text Similarity with Word Embeddings". In: *Proc. CIKM*. Vol. 15. 115.

Kim, Y. 2014. "Convolutional neural networks for sentence classification". *arXiv preprint arXiv:1408.5882*.

Kim, Y., Y. Jernite, D. Sontag, and A. M. Rush. 2015. "Character-aware neural language models". *arXiv preprint arXiv:1508.06615*.

Kingma, D. P. and M. Welling. 2014. "Auto-encoding variational bayes". In: *Proc. ICLR*.

Kiros, R., R. Zemel, and R. R. Salakhutdinov. 2014. "A multiplicative model for learning distributed text-based attribute representations". In: *Proc. NIPS*. 2348–2356.

Koch, G. 2015. "Siamese neural networks for one-shot image recognition". *PhD thesis*. University of Toronto.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. "Imagenet classification with deep convolutional neural networks". In: *Proc. NIPS*. 1097–1105.

Kusner, M., Y. Sun, N. Kolkin, and K. Weinberger. 2015. "From word embeddings to document distances". In: *Proc. ICML*. 957–966.

Lafferty, J. and C. Zhai. 2001. "Document language models, query models, and risk minimization for information retrieval". In: *Proc. SIGIR*. ACM. 111–119.

Lagun, D., C.-H. Hsieh, D. Webster, and V. Navalpakkam. 2014. "Towards better measurement of attention and satisfaction in mobile search". In: *Proc. SIGIR*. ACM. 113–122.

Larsson, G., M. Maire, and G. Shakhnarovich. 2016. "Fractalnet: Ultra-deep neural networks without residuals". *arXiv preprint arXiv:1605.07648*.

Lavrenko, V. 2008. *A generative theory of relevance*. Vol. 26. Springer Science & Business Media.

Lavrenko, V. and W. B. Croft. 2001. "Relevance based language models". In: *Proc. SIGIR*. ACM. 120–127.

Le, H.-S., I. Oparin, A. Allauzen, J.-L. Gauvain, and F. Yvon. 2011. "Structured output layer neural network language model". In: *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE. 5524–5527.

Le, Q. V. and T. Mikolov. 2014. "Distributed Representations of Sentences and Documents." In: *ICML*. Vol. 14. 1188–1196.

Lebret, R. and R. Collobert. 2013. "Word emdeddings through hellinger PCA". *arXiv preprint arXiv:1312.5542*.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep learning". *Nature*. 521(7553): 436–444.

LeCun, Y., F. J. Huang, and L. Bottou. 2004. "Learning methods for generic object recognition with invariance to pose and lighting". In: *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Vol. 2. IEEE. II–104.

LeCun, Y., K. Kavukcuoglu, and C. Farabet. 2010. "Convolutional networks and applications in vision". In: *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*. IEEE. 253–256.

Levy, O. and Y. Goldberg. 2014. "Dependencybased word embeddings". In: *Proc. ACL*. Vol. 2. 302–308.

Levy, O., Y. Goldberg, and I. Dagan. 2015. "Improving distributional similarity with lessons learned from word embeddings". *Transactions of the Association for Computational Linguistics*. 3: 211–225.

Levy, O., Y. Goldberg, and I. Ramat-Gan. 2014. "Linguistic regularities in sparse and explicit word representations". *CoNLL-2014*: 171.

Levy, S. 2011. *In the plex: How Google thinks, works, and shapes our lives*. Simon and Schuster.

Li, H. and Z. Lu. "Deep Learning for Information Retrieval".

Li, H., J. Xu, *et al.* 2014. “Semantic matching in search”. *Foundations and Trends® in Information Retrieval.* 7(5): 343–469.

Li, P., Q. Wu, and C. J. Burges. 2008. “Mcrank: Learning to rank using multiple classification and gradient boosting”. In: *Advances in neural information processing systems.* 897–904.

Liebling, D. J., P. N. Bennett, and R. W. White. 2012. “Anticipatory search: using context to initiate search”. In: *SIGIR.* ACM. 1035–1036.

Liu, C.-W., R. Lowe, I. V. Serban, M. Noseworthy, L. Charlin, and J. Pineau. 2016. “How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation”. *arXiv preprint arXiv:1603.08023.*

Liu, T.-Y. 2009. “Learning to Rank for Information Retrieval”. *Foundation and Trends in Information Retrieval.* 3(3): 225–331.

Liu, T.-Y., J. Xu, T. Qin, W. Xiong, and H. Li. 2007. “Letor: Benchmark dataset for research on learning to rank for information retrieval”. In: *Proceedings of SIGIR 2007 workshop on learning to rank for information retrieval.* 3–10.

Liu, X., J. Gao, X. He, L. Deng, K. Duh, and Y.-Y. Wang. “Representation Learning Using Multi-Task Deep Neural Networks for Semantic Classification and Information Retrieval”. *Proc. NAACL, May 2015.*

Lu, Z. and H. Li. 2013. “A deep architecture for matching short texts”. In: *Proc. NIPS.* 1367–1375.

Luce, R. D. 1959. *Individual Choice Behavior: A Theoretical Analysis.* John Wiley.

Lund, K. and C. Burgess. 1996. “Producing high-dimensional semantic spaces from lexical co-occurrence”. *Behavior Research Methods, Instruments, & Computers.* 28(2): 203–208.

Luong, M.-T., H. Pham, and C. D. Manning. 2015. “Effective approaches to attention-based neural machine translation”. *arXiv preprint arXiv:1508.04025.*

Luukkonen, P., M. Koskela, and P. Floréen. 2016. “Lstm-based predictions for proactive information retrieval”. *arXiv preprint arXiv:1606.06137.*

Ma, L., Z. Lu, L. Shang, and H. Li. 2015. "Multimodal convolutional neural networks for matching image and sentence". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2623–2631.

Maaten, L. v. d. and G. Hinton. 2008. "Visualizing data using t-SNE". *Journal of Machine Learning Research*. 9(Nov): 2579–2605.

MacKay, D. J. and L. C. B. Peto. 1995. "A hierarchical Dirichlet language model". *Natural language engineering*. 1(3): 289–308.

Mallows, C. L. 1957. "Non-null ranking models. I". *Biometrika*. 44(1/2): 114–130.

Manning, C. 2016. "Understanding Human Language: Can NLP and Deep Learning Help?" In: *Proc. SIGIR*. ACM. 1–1.

Manning, C. D., P. Raghavan, H. Schütze, *et al.* 2008. *Introduction to information retrieval*. Vol. 1. Cambridge university press Cambridge.

Markovsky, I. 2011. *Low rank approximation: algorithms, implementation, applications*. Springer Science & Business Media.

Matveeva, I., C. Burges, T. Burkard, A. Laucius, and L. Wong. 2006. "High accuracy retrieval with multiple nested ranker". In: *Proc. SIGIR*. ACM. 437–444.

Metzler, D. and W. B. Croft. 2005. "A Markov random field model for term dependencies". In: *Proc. SIGIR*. ACM. 472–479.

Metzler, D., S. Dumais, and C. Meek. 2007. "Similarity measures for short segments of text". In: *European Conference on Information Retrieval*. Springer. 16–27.

Miao, J., J. X. Huang, and Z. Ye. 2012. "Proximity-based rocchio's model for pseudo relevance". In: *Proc. SIGIR*. ACM. 535–544.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013a. "Efficient estimation of word representations in vector space". *arXiv preprint arXiv:1301.3781*.

Mikolov, T., M. Karafiát, L. Burget, J. Cernocky, and S. Khudanpur. 2010. "Recurrent neural network based language model." In: *Interspeech*. Vol. 2. 3.

Mikolov, T., S. Kombrink, L. Burget, J. Cernocky, and S. Khudanpur. 2011. "Extensions of recurrent neural network language model". In: *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE. 5528–5531.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013b. “Distributed representations of words and phrases and their compositionality”. In: *Proc. NIPS*. 3111–3119.

Mikolov, T., W.-t. Yih, and G. Zweig. 2013c. “Linguistic Regularities in Continuous Space Word Representations.” In: *HLT-NAACL*. Citeseer. 746–751.

Mitra, B. 2015. “Exploring Session Context using Distributed Representations of Queries and Reformulations”. In: *Proc. SIGIR*. ACM. 3–12.

Mitra, B. and N. Craswell. 2015. “Query Auto-Completion for Rare Prefixes”. In: *Proc. CIKM*. ACM.

Mitra, B. and N. Craswell. 2017. “Neural Text Embeddings for Information Retrieval”. In: *Proc. WSDM*. ACM. 813–814.

Mitra, B., F. Diaz, and N. Craswell. 2017a. “Learning to Match Using Local and Distributed Representations of Text for Web Search”. In: *Proc. WWW*. 1291–1299.

Mitra, B., F. Diaz, and N. Craswell. 2017b. “Luandri: a Clean Lua Interface to the Indri Search Engine”. In: *Proc. SIGIR*. ACM.

Mitra, B., E. Nalisnick, N. Craswell, and R. Caruana. 2016a. “A Dual Embedding Space Model for Document Ranking”. *arXiv preprint arXiv:1602.01137*.

Mitra, B., M. Shokouhi, F. Radlinski, and K. Hofmann. 2014. “On User Interactions with Query Auto-Completion”. In: *Proc. SIGIR*. 1055–1058.

Mitra, B., G. Simon, J. Gao, N. Craswell, and L. Deng. 2016b. “A Proposal for Evaluating Answer Distillation from Web Data”. In: *Proceedings of the SIGIR 2016 WebQA Workshop*.

Mnih, A. and G. E. Hinton. 2009. “A scalable hierarchical distributed language model”. In: *Advances in neural information processing systems*. 1081–1088.

Mnih, A. and Y. W. Teh. 2012. “A fast and simple algorithm for training neural probabilistic language models”. *arXiv preprint arXiv:1206.6426*.

Mnih, V., N. Heess, A. Graves, and K. Kavukcuoglu. 2014. “Recurrent models of visual attention”. In: *Proc. NIPS*. 2204–2212.

Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. 2015. "Human-level control through deep reinforcement learning". *Nature*. 518(7540): 529–533.

Montufar, G. F., R. Pascanu, K. Cho, and Y. Bengio. 2014. "On the number of linear regions of deep neural networks". In: *Proc. NIPS*. 2924–2932.

Morin, F. and Y. Bengio. 2005. "Hierarchical Probabilistic Neural Network Language Model." In: *Aistats*. Vol. 5. Citeseer. 246–252.

Nair, V. and G. E. Hinton. 2010. "Rectified linear units improve restricted boltzmann machines". In: *Proc. ICML*. 807–814.

Nalisnick, E., B. Mitra, N. Craswell, and R. Caruana. 2016. "Improving Document Ranking with Dual Word Embeddings". In: *Proc. WWW*.

Nanni, F., B. Mitra, M. Magnusson, and L. Dietz. 2017. "Benchmark for Complex Answer Retrieval". In: *Proc. ICTIR*. Amsterdam, The Netherlands: ACM.

Neubig, G., C. Dyer, Y. Goldberg, A. Matthews, W. Ammar, A. Anastasopoulos, M. Ballesteros, D. Chiang, D. Clothiaux, T. Cohn, *et al.* 2017. "DyNet: The Dynamic Neural Network Toolkit". *arXiv preprint arXiv:1701.03980*.

Nguyen, G.-H., L. Tamine, L. Soulier, and N. Bricon-Souf. 2016a. "Toward a deep neural approach for knowledge-based ir". *arXiv preprint arXiv:1606.07211*.

Nguyen, T., M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder, and L. Deng. 2016b. "MS MARCO: A Human Generated MAchine Reading COmprehension Dataset". *arXiv preprint arXiv:1611.09268*.

Nogueira, R. and K. Cho. 2017. "Task-Oriented Query Reformulation with Reinforcement Learning". In: *Proc. EMNLP*. 574–583.

Oard, D. W. and A. R. Diekema. 1998. "Cross-language information retrieval." *Annual Review of Information Science and Technology (ARIST)*. 33: 223–56.

Onal, K. D., Y. Zhang, I. S. Altingovde, M. M. Rahman, P. Karagoz, A. Braylan, B. Dang, H.-L. Chang, H. Kim, Q. McNamara, *et al.* 2018. "Neural information retrieval: At the end of the early years". *Information Retrieval Journal*. 21(2-3): 111–182.

Palangi, H., L. Deng, Y. Shen, J. Gao, X. He, J. Chen, X. Song, and R. Ward. 2014. "Semantic Modelling with Long-Short-Term Memory for Information Retrieval". *arXiv preprint arXiv:1412.6629*.

Palangi, H., L. Deng, Y. Shen, J. Gao, X. He, J. Chen, X. Song, and R. Ward. 2015. "Deep Sentence Embedding Using the Long Short Term Memory Network: Analysis and Application to Information Retrieval". *arXiv preprint arXiv:1502.06922*.

Pang, L., Y. Lan, J. Guo, J. Xu, and X. Cheng. 2016a. "A study of match-pyramid models on ad-hoc retrieval". *arXiv preprint arXiv:1606.04648*.

Pang, L., Y. Lan, J. Guo, J. Xu, S. Wan, and X. Cheng. 2016b. "Text Matching as Image Recognition". In: *Proc. AAAI*.

Pass, G., A. Chowdhury, and C. Torgeson. 2006. "A picture of search". In: *Proc. InfoScale*. ACM.

Pennington, J., R. Socher, and C. D. Manning. 2014. "Glove: Global vectors for word representation". *Proc. EMNLP*. 12: 1532–1543.

Plackett, R. L. 1975. "The analysis of permutations". *Applied Statistics*: 193–202.

Ponte, J. M. and W. B. Croft. 1998. "A language modeling approach to information retrieval". In: *Proc. SIGIR*. ACM. 275–281.

Rajaraman, A. 2008. "More data usually beats better algorithms". *Datawocky Blog*.

Rajpurkar, P., J. Zhang, K. Lopyrev, and P. Liang. 2016. "Squad: 100,000+ questions for machine comprehension of text". *arXiv preprint arXiv:1606.05250*.

Ranzato, M., C. Poultney, S. Chopra, and Y. LeCun. 2006. "Efficient learning of sparse representations with an energy-based model". In: *Proc. NIPS*. MIT Press. 1137–1144.

Rezende, D. J., S. Mohamed, and D. Wierstra. 2014. "Stochastic back-propagation and approximate inference in deep generative models". In: *Proc. ICML*.

Al-Rfou, R., G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky, *et al.* 2016. "Theano: A Python framework for fast computation of mathematical expressions". *arXiv preprint arXiv:1605.02688*.

Richardson, M., C. J. Burges, and E. Renshaw. 2013. "MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text." In: *EMNLP*. Vol. 3. 4.

Robertson, S. E., E. Kanoulas, and E. Yilmaz. 2010. "Extending average precision to graded relevance judgments". In: *Proc. SIGIR*. ACM. 603–610.

Robertson, S. E., S. Walker, M. Beaulieu, M. Gatford, and A. Payne. 1996. "Okapi at TREC-4". In: *Proc. TREC*. Vol. 500. 73–97.

Robertson, S. and H. Zaragoza. 2009. "The probabilistic relevance framework: BM25 and beyond". *Foundations and Trends® in Information Retrieval*. 3(4): 333–389.

Robertson, S., H. Zaragoza, and M. Taylor. 2004. "Simple BM25 extension to multiple weighted fields". In: *Proc. CIKM*. ACM. 42–49.

Rohde, D. L., L. M. Gonnerman, and D. C. Plaut. 2006. "An improved model of semantic similarity based on lexical co-occurrence". *Communications of the ACM*. 8: 627–633.

Rong, X. 2014. "word2vec Parameter Learning Explained". *arXiv preprint arXiv:1411.2738*.

Rosset, C., D. Jose, G. Ghosh, B. Mitra, and S. Tiwary. 2018. "Optimizing Query Evaluations using Reinforcement Learning for Web Search". In: *Proc. SIGIR*. ACM.

Roy, D., D. Paul, M. Mitra, and U. Garain. 2016. "Using Word Embeddings for Automatic Query Expansion". *arXiv preprint arXiv:1606.07608*.

Rubner, Y., C. Tomasi, and L. J. Guibas. 1998. "A metric for distributions with applications to image databases". In: *Computer Vision, 1998. Sixth International Conference on*. IEEE. 59–66.

Ruder, S. 2016. "Approximating the Softmax for Learning Word Embeddings". Accessed June 7, 2017. URL: http://sebastianruder.com/word-embeddings-softmax/.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. 2015. "ImageNet Large Scale Visual Recognition Challenge". *International Journal of Computer Vision (IJCV)*. 115(3): 211–252. DOI: 10.1007/s11263-015-0816-y.

Sahlgren, M. 2006. "The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces". *PhD thesis*. Institutionen för lingvistik.

Sak, H., A. W. Senior, and F. Beaufays. 2014. "Long short-term memory recurrent neural network architectures for large scale acoustic modeling." In: *Interspeech*. 338–342.

Salakhutdinov, R. and G. Hinton. 2009. "Semantic hashing". *International Journal of Approximate Reasoning*. 50(7): 969–978.

Schmidhuber, J. 2015. "Deep learning in neural networks: An overview". *Neural networks*. 61: 85–117.

Schroff, F., D. Kalenichenko, and J. Philbin. 2015. "Facenet: A unified embedding for face recognition and clustering". In: *Proc. CVPR*. 815–823.

Senécal, J.-S. and Y. Bengio. 2003. "Adaptive importance sampling to accelerate training of a neural probabilistic language model". *Tech. rep.* Technical report, IDIAP.

Severyn, A. and A. Moschitti. 2015. "Learning to rank short text pairs with convolutional deep neural networks". In: *Proc. SIGIR*. ACM. 373–382.

Shan, Y., T. R. Hoens, J. Jiao, H. Wang, D. Yu, and J. Mao. 2016. "Deep Crossing: Web-scale modeling without manually crafted combinatorial features". In: *Proc. SIGKDD*. ACM. 255–262.

Shen, Y., X. He, J. Gao, L. Deng, and G. Mesnil. 2014a. "A latent semantic model with convolutional-pooling structure for information retrieval". In: *Proc. CIKM*. ACM. 101–110.

Shen, Y., X. He, J. Gao, L. Deng, and G. Mesnil. 2014b. "Learning semantic representations using convolutional neural networks for Web search". In: *Proc. WWW*. 373–374.

Shi, S., Q. Wang, P. Xu, and X. Chu. 2016. "Benchmarking State-of-the-Art Deep Learning Software Tools". *arXiv preprint arXiv:1608.07249*.

Shokouhi, M. and Q. Guo. 2015. "From queries to cards: Re-ranking proactive card recommendations based on reactive search history". In: *SIGIR*. ACM. 695–704.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.* 2016. "Mastering the game of Go with deep neural networks and tree search". *Nature*. 529(7587): 484–489.

Singhal, A., C. Buckley, and M. Mitra. 1996. "Pivoted document length normalization". In: *Proc. SIGIR*. ACM. 21–29.

Socher, R., C. C. Lin, C. Manning, and A. Y. Ng. 2011a. "Parsing natural scenes and natural language with recursive neural networks". In: *Proc. ICML*. 129–136.

Socher, R., J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning. 2011b. "Semi-supervised recursive autoencoders for predicting sentiment distributions". In: *Proc. EMNLP*. Association for Computational Linguistics. 151–161.

Song, Y. and Q. Guo. 2016. "Query-less: Predicting task repetition for nextgen proactive search and recommendation engines". In: *WWW*. International World Wide Web Conferences Steering Committee. 543–553.

Sordoni, A., Y. Bengio, H. Vahabi, C. Lioma, J. G. Simonsen, and J.-Y. Nie. 2015a. "A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion". *arXiv preprint arXiv:1507.02221*.

Sordoni, A., M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J.-Y. Nie, J. Gao, and B. Dolan. 2015b. "A neural network approach to context-sensitive generation of conversational responses". In: *Proceedings of NAACL-HLT*. arXiv:1506.06714. 196–205.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. "Dropout: A simple way to prevent neural networks from overfitting". *The Journal of Machine Learning Research*. 15(1): 1929–1958.

Strohman, T., D. Metzler, H. Turtle, and W. B. Croft. 2005. "Indri: A language model-based search engine for complex queries". In: *Proceedings of the International Conference on Intelligent Analysis*. Vol. 2. No. 6. Citeseer. 2–6.

Sturm, B. L. 2014. “A simple method to determine if a music information retrieval system is a “horse””. *IEEE Transactions on Multimedia*. 16(6): 1636–1644.

Sukhbaatar, S., J. Weston, R. Fergus, *et al.* 2015. “End-to-end memory networks”. In: *Proc. NIPS*. 2440–2448.

Sun, C., A. Shrivastava, S. Singh, and A. Gupta. 2017. “Revisiting unreasonable effectiveness of data in deep learning era”. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE. 843–852.

Sun, F., J. Guo, Y. Lan, J. Xu, and X. Cheng. 2015. “Learning word representations by jointly modeling syntagmatic and paradigmatic relations”. In: *Proc. ACL*.

Sun, F., J. Guo, Y. Lan, J. Xu, and X. Cheng. 2016a. “Semantic Regularities in Document Representations”. *arXiv preprint arXiv:1603.07603*.

Sun, F., J. Guo, Y. Lan, J. Xu, and X. Cheng. 2016b. “Sparse word embeddings using l1 regularized online learning”. In: *Proc. IJCAI*. 2915–2921.

Sutskever, I., J. Martens, and G. E. Hinton. 2011. “Generating text with recurrent neural networks”. In: *Proc. ICML*. 1017–1024.

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. 2015. “Going deeper with convolutions”. In: *Proc. CVPR*. 1–9.

Szegedy, C., W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. 2013. “Intriguing properties of neural networks”. *arXiv preprint arXiv:1312.6199*.

Tai, K. S., R. Socher, and C. D. Manning. 2015. “Improved semantic representations from tree-structured long short-term memory networks”. *arXiv preprint arXiv:1503.00075*.

Taylor, M., H. Zaragoza, N. Craswell, S. Robertson, and C. Burges. 2006. “Optimisation methods for ranking functions with multiple parameters”. In: *Proc. CIKM*. ACM. 585–593.

Tishby, N., F. C. Pereira, and W. Bialek. 2000. “The information bottleneck method”. *arXiv preprint physics/0004057*.

Tokui, S., K. Oono, S. Hido, and J. Clayton. 2015. "Chainer: a next-generation open source framework for deep learning". In: *Proceedings of workshop on machine learning systems (LearningSys) in the twenty-ninth annual conference on neural information processing systems (NIPS)*.

Turney, P. D. and P. Pantel. 2010. "From frequency to meaning: Vector space models of semantics". *Journal of artificial intelligence research*. 37: 141–188.

Van Gysel, C., E. Kanoulas, and M. de Rijke. 2017a. "Pyndri: a Python Interface to the Indri Search Engine". *arXiv preprint arXiv:1701.00749*.

Van Gysel, C., B. Mitra, M. Venanzi, R. Rosemarin, G. Kukla, P. Grudzien, and N. Cancedda. 2017b. "Reply With: Proactive Recommendation of Email Attachments". In: *Proc. CIKM*.

Van Gysel, C., M. de Rijke, and E. Kanoulas. 2016a. "Learning latent vector spaces for product search". In: *Proc. CIKM*. ACM. 165–174.

Van Gysel, C., M. de Rijke, and M. Worring. 2016b. "Unsupervised, efficient and semantic expertise retrieval". In: *Proc. WWW*. International World Wide Web Conferences Steering Committee. 1069–1079.

Varma, M. and M. Cisse. 2015. "Extreme Classification 2015: The NIPS Workshop on Multi-class and Multi-label Learning in Extremely Large Label Spaces". Accessed June 7, 2017. URL: http://manikvarma.org/events/XC15/.

Vaswani, A., Y. Zhao, V. Fossum, and D. Chiang. 2013. "Decoding with Large-Scale Neural Language Models Improves Translation." In: *EMNLP*. 1387–1392.

Vincent, P., A. de Brébisson, and X. Bouthillier. 2016. "Exact gradient updates in time independent of output size for the spherical loss family". *arXiv preprint arXiv:1606.08061*.

Vinyals, O., C. Blundell, T. Lillicrap, D. Wierstra, *et al.* 2016. "Matching networks for one shot learning". In: *Proc. NIPS*. 3630–3638.

Vinyals, O. and Q. Le. 2015. "A neural conversational model". *ICML Deep Learning Workshop*. arXiv:1506.05869.

Voorhees, E. M. and D. Harman. 2000. "Overview of the eighth text retrieval conference (TREC-8)": 1–24.

Voorhees, E. M., D. K. Harman, *et al.* 2005. *TREC: Experiment and evaluation in information retrieval.* Vol. 1. MIT press Cambridge.

Voorhees, E. M. and D. M. Tice. 2000. "Building a question answering test collection". In: *Proc. SIGIR*. ACM. 200–207.

Vulić, I. and M.-F. Moens. 2015. "Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings". In: *Proc. SIGIR*. ACM. 363–372.

Wan, S., Y. Lan, J. Guo, J. Xu, L. Pang, and X. Cheng. 2015. "A deep architecture for semantic matching with multiple positional sentence representations". *arXiv preprint arXiv:1511.08277*.

Wan, S., Y. Lan, J. Xu, J. Guo, L. Pang, and X. Cheng. 2016. "Match-srnn: Modeling the recursive matching structure with spatial rnn". *arXiv preprint arXiv:1604.04378*.

Wan, X. 2007. "A novel document similarity measure based on earth mover's distance". *Information Sciences*. 177(18): 3718–3730.

Wan, X. and Y. Peng. 2005. "The earth mover's distance as a semantic measure for document similarity". In: *Proc. CIKM*. ACM. 301–302.

Wang, J., L. Yu, W. Zhang, Y. Gong, Y. Xu, B. Wang, P. Zhang, and D. Zhang. 2017. "IRGAN: A Minimax Game for Unifying Generative and Discriminative Information Retrieval Models". *arXiv preprint arXiv:1705.10513*.

Wang, L., J. Lin, and D. Metzler. 2011. "A cascade ranking model for efficient ranked retrieval". In: *Proc. SIGIR*. ACM. 105–114.

Weston, J., A. Bordes, S. Chopra, A. M. Rush, B. van Merriënboer, A. Joulin, and T. Mikolov. 2015. "Towards ai-complete question answering: A set of prerequisite toy tasks". *arXiv preprint arXiv:1502.05698*.

Weston, J., S. Chopra, and A. Bordes. 2014. "Memory networks". *arXiv preprint arXiv:1410.3916*.

White, R. W. 2016. *Interactions with search systems*. Cambridge University Press.

Wissner-Gross, A. 2016. "Datasets Over Algorithms". URL: https://www.edge.org/response-detail/26587.

Wu, Q., C. J. Burges, K. M. Svore, and J. Gao. 2010. "Adapting boosting for information retrieval measures". *Information Retrieval*. 13(3): 254–270.

Xia, F., T.-Y. Liu, J. Wang, W. Zhang, and H. Li. 2008. "Listwise approach to learning to rank: theory and algorithm". In: *Proc. ICML*. ACM. 1192–1199.

Xia, L., J. Xu, Y. Lan, J. Guo, and X. Cheng. 2016. "Modeling Document Novelty with Neural Tensor Network for Search Result Diversification". In: *Proc. SIGIR*. ACM. 395–404.

Xie, Y. and D. O'Hallaron. 2002. "Locality in search engine queries and its implications for caching". In: *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*. Vol. 3. IEEE. 1238–1247.

Xu, K., J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio. 2015. "Show, attend and tell: Neural image caption generation with visual attention". In: *International Conference on Machine Learning*. 2048–2057.

Yan, R., Y. Song, and H. Wu. 2016. "Learning to respond with deep neural networks for retrieval-based human-computer conversation system". In: *Proc. SIGIR*. ACM. 55–64.

Yan, X., J. Guo, S. Liu, X. Cheng, and Y. Wang. 2013. "Learning topics in short texts by non-negative matrix factorization on term correlation matrix". In: *Proceedings of the SIAM International Conference on Data Mining*.

Yang, L., Q. Ai, J. Guo, and W. B. Croft. 2016. "aNMM: Ranking Short Answer Texts with Attention-Based Neural Matching Model". In: *Proc. CIKM*. ACM. 287–296.

Yang, Y., W.-t. Yih, and C. Meek. 2015. "WikiQA: A Challenge Dataset for Open-Domain Question Answering." In: *EMNLP*. Citeseer. 2013–2018.

Yih, W.-t., K. Toutanova, J. C. Platt, and C. Meek. 2011. "Learning discriminative projections for text similarity measures". In: *Proc. CoNLL*. Association for Computational Linguistics. 247–256.

Yin, W., H. Schütze, B. Xiang, and B. Zhou. 2015. "Abcnn: Attention-based convolutional neural network for modeling sentence pairs". *arXiv preprint arXiv:1512.05193*.

Yu, D., A. Eversole, M. Seltzer, K. Yao, Z. Huang, B. Guenter, O. Kuchaiev, Y. Zhang, F. Seide, H. Wang, *et al.* "An introduction to computational networks and the computational network toolkit".

Yue, Y. and C. Burges. 2007. "On using simultaneous perturbation stochastic approximation for IR measures, and the empirical optimality of LambdaRank". In: *NIPS Machine Learning for Web Search Workshop*.

Yue, Y., T. Finley, F. Radlinski, and T. Joachims. 2007. "A support vector method for optimizing average precision". In: *Proc. SIGIR*. ACM. 271–278.

Zamani, H. and W. B. Croft. 2016a. "Embedding-based query language models". In: *Proc. ICTIR*. ACM. 147–156.

Zamani, H. and W. B. Croft. 2016b. "Estimating embedding vectors for queries". In: *Proc. ICTIR*. ACM. 123–132.

Zamani, H., B. Mitra, X. Song, N. Craswell, and S. Tiwary. 2018. "Neural Ranking Models with Multiple Document Fields". In: *Proc. WSDM*.

Zaragoza, H., N. Craswell, M. J. Taylor, S. Saria, and S. E. Robertson. 2004. "Microsoft Cambridge at TREC 13: Web and Hard Tracks." In: *TREC*. Vol. 4. 1–1.

Zhai, C. and J. Lafferty. 2001. "A study of smoothing methods for language models applied to ad hoc information retrieval". In: *Proc. SIGIR*. ACM. 334–342.

Zhang, C., S. Bengio, M. Hardt, B. Recht, and O. Vinyals. 2016. "Understanding deep learning requires rethinking generalization". *arXiv preprint arXiv:1611.03530*.

Zhao, H., Z. Lu, and P. Poupart. 2015. "Self-Adaptive Hierarchical Sentence Model". *arXiv preprint arXiv:1504.05070*.

Zheng, G. and J. Callan. 2015. "Learning to Reweight Terms with Distributed Representations". In: *Proc. SIGIR*. ACM. 575–584.

Zhou, X., D. Dong, H. Wu, S. Zhao, R. Yan, D. Yu, X. Liu, and H. Tian. 2016. "Multi-view response selection for human-computer conversation". *EMNLP'16*.

Zhu, M. 2004. "Recall, precision and average precision". *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo*. 2: 30.

Zoph, B., A. Vaswani, J. May, and K. Knight. 2016. "Simple, fast noise-contrastive estimation for large rnn vocabularies". In: *Proceedings of NAACL-HLT*. 1217–1222.

Zuccon, G., B. Koopman, P. Bruza, and L. Azzopardi. 2015. "Integrating and evaluating neural word embeddings in information retrieval". In: *Proc. ADCS*. ACM. 12.

Zweig, G. and K. Makarychev. 2013. "Speed regularization and optimality in word classing". In: *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE. 8237–8241.