Fusion in Information Retrieval

SIGIR 2018 Half-Day Tutorial

Oren Kurland
Technion — Israel Institute of Technology
Haifa, Israel
kurland@ie.technion.ac.il

J. Shane Culpepper
RMIT University
Melbourne, Australia
shane.culpepper@rmit.edu.au

ABSTRACT

Fusion is an important and central concept in Information Retrieval. The goal of fusion methods is to merge different sources of information so as to address a retrieval task. For example, in the adhoc retrieval setting, fusion methods have been applied to merge multiple document lists retrieved for a query. The lists could be retrieved using different query representations, document representations, ranking functions and corpora. The goal of this half day, intermediate-level, tutorial is to provide a methodological view of the theoretical foundations of fusion approaches, the numerous fusion methods that have been devised and a variety of applications for which fusion techniques have been applied.

1 MOTIVATION

Fusion is a classic technique used for more than twenty years in Information Retrieval, specifically adhoc (query-based) retrieval, that allows multiple sources of information to be combined into a single result set [32, 40]. Fusion can be collection-based, system-based (multiple ranking algorithms), content-based, and even query-based when many similar queries express the same information need [32]. The real power of fusion comes from the fact that even simple aggregation functions have the potential to provide enhanced retrieval effectiveness by exploiting the *chorus effect* [96].

In this tutorial, we will show that advances in fusion are directly applicable to current open problems in the Information Retrieval community, and that much can be learned from these models as machine learning becomes even more prominent in modern search solutions. In particular we draw parallels between unsupervised fusion and ensembles of classifiers in supervised learning [36, 82, 116].

We focus on retrieval settings where a single corpus is used, and different factors that affect retrieval vary; e.g., queries used to represent the information need, document and/or query representations, ranking functions, etc. We briefly discuss the setting of retrieval over several corpora (a.k.a., federated or distributed search [24, 90]); specifically, we survey several state-of-the-art techniques for fusing lists retrieved from different corpora. We believe that federated search deserves a tutorial in its own right which covers the three main challenges: resource representation, resource selection and results merging [24, 47, 90].

Table 1: Effectiveness comparison of three state-of-the-art ranking methods for the most common query variation for each topic from the ClueWeb12B UQV100 collection [10]. Here ‡ means $p < 0.001$ in a Bonferroni corrected two-tailed t-test.

Method	NDCG@10	W/T/L
BM25	0.212	—/—/—
SDM-Field	0.233	57/3/40
LambdaMART	0.225	59/2/39
DoubleFuse, v=all	$0.300^\ddagger$	80/1/19

Finally, it is important for everyone in the community to understand just how effective simple fusion techniques can be. Figure 1 and Table 1 compare three state-of-the-art retrieval systems on 100 adhoc queries in the ClueWeb12B UQV100 collection. The three systems being compared are BM25, a field-based SDM model [76] (the exact configuration is identical to the one described by Gallagher et al. [42]), a LambdaMART learning-to-rank (LTR) model [23, 26] (here lightGBM is used with 459 features), and double unsupervised fusion [11, 18] (RRF [29] over all UQV query variations and two systems - SDM-Field and BM25). Figure 1 shows the three strong baselines as a difference in NDCG@10 score w.r.t. a BM25 bag-of-words run. We can clearly see that not only does fusion make more queries better on average, as shown in Table 1, it is also far less likely to make queries *worse*. This can clearly be seen when comparing Wins, Ties, and Losses (W/T/L) in the table. So, there is much to be learned from fusion baselines when doing exploratory failure analysis on the robustness of new ranking algorithms.

2 TUTORIAL OBJECTIVES

- Highlight the important role of fusion in Information Retrieval.
- Provide a methodological view of the numerous fusion methods.
- Provide an overview of the theoretical foundations of various fusion approaches.
- Introduce the audience to various tasks and challenges for which fusion has been applied and can be applied.
- Discuss parallels with, and more generally pointers to, relevant, related work in machine learning and computational social choice theory.
- Discuss open questions and challenges.

SIGIR '18, Ann Arbor, MI, USA
 978-1-4503-5657-2/18/07…$15.00
DOI: 10.1145/3209978.3210186

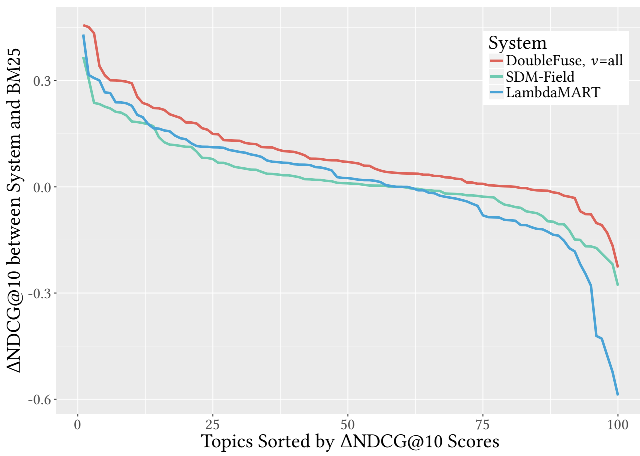

Figure 1: Per topic breakdown comparison of NDCG@10 differences of several state-of-the-art adhoc ranking techniques. The scores shown are the difference between the method and a simple BM25 bag-of-words run. The Double Fusion Technique uses all of the query variations (v=all) for each of the 100 topics, uses RRF Fusion, and combines two systems – SDM-Field and BM25.

3 FORMAT AND PLANNED SCHEDULE

Table 2: Half Day Schedule of Topics

Time	Topic
9:00 - 9:15	Introduction
9:15 - 9:30	Historical Context
9:30 - 10:00	Theoretical Foundations
10:00 - 10:30	Fusion in Practice
10:30 - 11:00	Coffee Break
11:00 - 11:20	Fusion in Practice (contd.)
11:20 - 11:45	Learning & Fusion
11:45 - 12:10	Applications
12:10 - 12:30	Conclusions & Future Directions

OUTLINE

- Intro and Overview
- Historical Context
  - Social Choice Theory and Voting Schemes [20]
    - Condorcet, Borda, Kemeny [13, 34, 115]
  - TREC and Rank Fusion [40]
  - Federated Search [24, 90]
- Theoretical Foundations
  - The Fusion Hypothesis [14, 31, 32, 56, 57, 81, 94]
  - Classifier Combination [93]
  - Fusion Frameworks [3, 53, 55, 88, 96, 99, 100, 102]
- Fusion in Practice
  - Score-based (e.g., [3, 40, 56, 57, 78])
  - Rank-based (e.g., [11, 29, 38, 41, 77, 79, 80, 103])
  - Retrieval Score Normalization and Rank-to-Score Transformations [4, 5, 29, 39, 57, 73, 74, 78, 104, 108]
  - Content-based [15, 30, 49–51, 62, 64, 87, 91]
  - Selecting Retrieved Lists for Fusion [43, 45, 46]
  - Query Variations [11, 16–18, 22, 28, 52, 113]
  - Failure Analysis / Risk [18, 37]
  - Efficiency Considerations [44, 59]
- Learning & Fusion [55, 88]
  - Models over Permutations (e.g., [1, 38, 48, 54, 83])
  - Supervised (e.g., [3, 55, 65–67, 85, 88, 89, 102, 105, 106, 110, 112]) vs Unsupervised (e.g., [6, 9, 29, 40, 107])
  - Ensembles [36, 82, 116]
- Applications
  - Query Performance Prediction [7, 35, 75, 81, 85, 92, 95, 111]
  - Diversification [60, 63, 109]
  - Relevance Feedback [8, 84]
  - Selecting a Ranker [2, 12, 33, 58]
  - Blog and Microblog Retrieval [60, 61, 64, 101]
  - Pooling and Evaluation [8, 21, 25, 68–71, 86, 97, 98]
- Conclusions & Future Directions

4 TYPE OF SUPPORT MATERIALS TO BE SUPPLIED TO ATTENDEES

- A Web page that contains all materials.
- Downloadable slides available in PDF format.
- Extensive bibliography that helps to further explore topics discussed in the tutorial.
- Scripts and source code for common fusion techniques that can be used by PhD students in future work[1].

5 PRESENTERS' BIOGRAPHY

Oren Kurland is an Associate Professor at the Technion — Israel Institute of Technology. He holds a Ph.D. in Computer Science from Cornell University. Oren has served as a senior program committee member and/or area chair for the SIGIR, CIKM, WSDM, WWW and ECIR conferences for the last few years. He is also a member of the editorial board of the Information Retrieval Journal. Oren served as a doctoral consortium co-chair for WSDM 2014, and as a program co-chair for the ICTIR 2013 and SPIRE 2013 conferences. He has also served as the chair of the steering committee of the ACM SIGIR ICTIR conference.

Shane Culpepper is a Vice-Chancellor's Principal Research Fellow and Associate Professor at RMIT University, and is the Director of the Centre for Information Discovery. His research focuses on building next generation search engines, and exploring new ways to evaluate the quality of search. Research interests include information retrieval, text indexing, data compression, system evaluation, knowledge discovery, machine learning, natural language processing, algorithm engineering and scalability. He is active in many research capacities in the IR research community, including being on the editorial board for the Information Retrieval Journal, and routinely serves on the program committees at ADCS, CIKM, ICDE, SIGIR, SPIRE, WSDM, and WWW. He has been a Program Co-Chair for ADCS, the SIGIR Doctoral Consortium, and CIKM, and will be a General Chair for WSDM 2019. .

[1] https://www.github.com/rmit-ir/polyfuse

Acknowledgements. We thank the reviewers for their comments. This work was supported in part by the Australian Research Council's *Discovery Projects* Scheme (DP170102231) and a grant from the Mozilla Foundation.

REFERENCES

[1] N. Ailon. 2010. Aggregation of Partial Rankings, p-Ratings and Top-m Lists. *Algorithmica* 57, 2 (2010), 284–300.

[2] G. Amati, C. Carpineto, and G. Romano. 2004. Query difficulty, robustness, and selective application of query expansion. In *Proc. SIGIR*. 127–137.

[3] Y. Anava, A. Shtok, O. Kurland, and E. Rabinovich. 2016. A Probabilistic Fusion Framework. In *Proc. CIKM*. 1463–1472.

[4] A. Arampatzis and J. Kamps. 2009. A signal-to-noise approach to score normalization. In *Proc. CIKM*. 797–806.

[5] A. Arampatzis and S. Robertson. 2011. Modeling score distributions in information retrieval. *Inf. Retr.* 14, 1 (2011), 26–46.

[6] J. A. Aslam and M. Montague. 2001. Models for metasearch. In *Proc. SIGIR*. 276–284.

[7] J. A. Aslam and V. Pavlu. 2007. Query Hardness Estimation Using Jensen-Shannon Divergence Among Multiple Scoring Functions. In *Proc. ECIR*. 198–209.

[8] J. A. Aslam, V. Pavlu, and R. Savell. 2003. A unified model for metasearch and the efficient evaluation of retrieval systems via the hedge algorithm. In *Proc. SIGIR*. 393–394.

[9] J. A. Aslam, V. Pavlu, and E. Yilmaz. 2005. Measure-based Metasearch. In *Proc. SIGIR*. 571–572.

[10] P. Bailey, A. Moffat, F. Scholer, and P. Thomas. 2016. UQV100: A test collection with query variability. In *Proc. SIGIR*. 725–728.

[11] P. Bailey, A. Moffat, F. Scholer, and P. Thomas. 2017. Retrieval consistency in the presence of query variations. In *Proc. SIGIR*. 395–404.

[12] N. Balasubramanian and J. Allan. 2010. Learning to select rankers. In *Proc. SIGIR*. 855–856.

[13] J. Bartholdi, C. A. Tovey, and M. A. Trick. 1989. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare* 6, 2 (1989), 157–165.

[14] S. M. Beitzel, E. C. Jensen, A. Chowdhury, O. Frieder, D. A. Grossman, and N. Goharian. 2003. Disproving the Fusion Hypothesis: An Analysis of Data Fusion via Effective Information Retrieval Strategies. In *Proc. SAC*. 823–827.

[15] S. M. Beitzel, E. C. Jensen, O. Frieder, A. Chowdhury, and G. Pass. 2005. Surrogate scoring for improved metasearch precision. In *Proc. SIGIR*. 583–584.

[16] N. J. Belkin, C. Cool, W. B. Croft, and J. P. Callan. 1993. The Effect of Multiple Query Variations on Information Retrieval System Performance. In *Proc. SIGIR*. 339–346.

[17] N. J. Belkin, P. Kantor, E. A. Fox, and J. A. Shaw. 1995. Combining the evidence of multiple query representations for information retrieval. *Inf. Proc. & Man.* 31, 3 (1995), 431–448.

[18] R. Benham and J. S. Culpepper. 2017. Risk-reward Trade-offs in Rank Fusion. In *Proc. ADCS*. Article 1, 1:1–1:8 pages.

[19] R. Benham, L. Gallagher, J. Mackenzie, T. T. Damessie, R.-C. Chen, F. Scholer, A. Moffat, and J. S. Culpepper. 2017. RMIT at the TREC 2017 CORE Track.. In *Proc. TREC*.

[20] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia (Eds.). 2016. *Handbook of Computational Social Choice*. Cambridge University Press.

[21] C. Buckley, D. Dimmick, I. Soboroff, and E. M. Voorhees. 2007. Bias and the limits of pooling for large collections. *Inf. Retr.* (2007), 491–508.

[22] C. Buckley and J. Walz. 1999. The TREC-8 query track. In *Proc. TREC*.

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

[24] J. Callan. 2000. Distributed information retrieval. In *Advances in information retrieval*, W.B. Croft (Ed.). Kluwer Academic Publishers, Chapter 5, 127–150.

[25] B. Carterette, V. Pavlu, E. Kanoulas, J. A. Aslam, and J. Allan. 2008. Evaluation over thousands of queries. In *Proc. SIGIR*. 651–658.

[26] R.-C. Chen, L. Gallagher, R. Blanco, and J. S. Culpepper. 2017. Efficient cost-aware cascade ranking in multi-stage retrieval. In *Proc. SIGIR*. 445–454.

[27] F. M. Choudhury, Z. Bao, J. S. Culpepper, and T. Sellis. 2017. Monitoring the Top-m Rank Aggregation of Spatial Objects in Streaming Queries. In *Proc. ICDE*. 585–596.

[28] K. Collins-Thompson and J. Callan. 2007. Estimation and use of uncertainty in pseudo-relevance feedback. In *Proc. SIGIR*. 303–310.

[29] G. V. Cormack, C. L. A. Clarke, and S. Büttcher. 2009. Reciprocal rank fusion outperforms Condorcet and individual rank learning methods. In *Proc. SIGIR*. 758–759.

[30] N. Craswell, D. Hawking, and P. B. Thistlewaite. 1999. Merging results from isolated search engines. In *Proc. ADC*. 189–200.

[31] W. B. Croft (Ed.). 2000. *Advances in Information Retrieval: Recent Research from the Center for Intelligent Information Retrieval*. Number 7 in The Kluwer International Series on Information Retrieval. Kluwer.

[32] W. B. Croft. 2000. Combining approaches to information retrieval. See [31], Chapter 1, 1–36.

[33] S. Cronen-Townsend, Y. Zhou, and W. B. Croft. 2004. *A Language Modeling Framework for Selective Query Expansion*. Technical Report IR-338. Center for Intelligent Information Retrieval, University of Massachusetts.

[34] J. C. de Borda. 1784. Mémoire sur les élections au scrutin. *Histoire de l'Academie Royale des Sciences pour 1781 (Paris, 1784)* (1784).

[35] F. Diaz. 2007. Regularizing query-based retrieval scores. *Inf. Retr.* 10, 6 (2007), 531–562.

[36] T G Dietterich. 2000. Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems*. Springer, 1–15.

[37] B. T. Dinçer, C. Macdonald, and I. Ounis. 2014. Hypothesis testing for the risk-sensitive evaluation of retrieval systems. In *Proc. SIGIR*. 23–32.

[38] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. 2001. Rank Aggregation Methods for the Web. In *Proc. WWW*. 613–622.

[39] M. Efron. 2009. Generative model-based metasearch for data fusion in information retrieval. In *Proc. JCDL*. 153–162.

[40] E. A. Fox and J. A. Shaw. 1994. Combination of multiple searches. In *Proc. TREC*.

[41] H. D. Frank and I. Taksa. 2005. Comparing rank and score combination methods for data fusion in information retrieval. *Inf. Retr.* 8, 3 (2005), 449–480.

[42] L. Gallagher, J. Mackenzie, R. Benham, R.-C. Chen, F. Scholer, and J. S. Culpepper. 2017. RMIT at the NTCIR-13 We Want Web task. In *Proc. NTCIR*.

[43] N. P. Gopalan and K. Batri. 2007. Adaptive Selection of Top-m Retrieval Strategies for Data Fusion in Information Retrieval. *Intl. J. of Soft Computing* 2, 1 (2007).

[44] S. Huo, M. Zhang, Y. Liu, and S. Ma. 2014. Improving tail query performance by fusion model. In *Proc. CIKM*. 559–658.

[45] A. Juárez-González, M. Montes-y-Gómez, L. V. Pineda, and D. O. Arroyo. 2009. On the Selection of the Best Retrieval Result Per Query - An Alternative Approach to Data Fusion. In *Proc. FQAS*. 111–121.

[46] A. Juárez-González, M. Montes-y-Gómez, L. V. Pineda, D. P. Avendaño, and M. A. Pérez-Coutiño. 2010. Selecting the N-Top Retrieval Result Lists for an Effective Data Fusion. In *Proc. CICLing*. 580–589.

[47] Y. Kim, J. Callan, J. S. Culpepper, and A. Moffat. 2017. Efficient distributed selective search. *Inf. Retr.* 20, 3 (2017), 221–252.

[48] A. Klementiev, D. Roth, and K. Small. 2008. Unsupervised rank aggregation with distance-based models. In *Proc. ICML*. 472–479.

[49] A. K. Kozorovitzky and O. Kurland. 2009. From "Identical" to "Similar": Fusing Retrieved Lists Based on Inter-document Similarities. In *Proc. ICTIR*. 212–223.

[50] A. K. Kozorovitzky and O. Kurland. 2011. Cluster-based fusion of retrieved lists. In *Proc. SIGIR*. 893–902.

[51] A. K. Kozorovitzky and O. Kurland. 2011. From "Identical" to "Similar": Fusing Retrieved Lists Based on Inter-document Similarities. *J. of AI Res.* 41 (2011).

[52] K.-L. Kwok, L. Grunfeld, and P. Deng. 2005. Improving weak ad-hoc retrieval by web assistance and data fusion. In *Proc. AIRS*. 17–30.

[53] M. Lalmas. 2002. A Formal Model for Data Fusion. In *Proc. FQAS*. 274–288.

[54] G. Lebanon and J. D. Lafferty. 2002. Cranking: Combining Rankings Using Conditional Probability Models on Permutations. In *Proc. ICML*. 363–370.

[55] C.-J. Lee, Q. Ai, W. B. Croft, and D. Sheldon. 2015. An Optimization Framework for Merging Multiple Result Lists. In *Proc. CIKM*. 303–312.

[56] J. H. Lee. 1995. Combining multiple evidence from different properties of weighting schemes. In *Proc. SIGIR*. 180–188.

[57] J. H. Lee. 1997. Analyses of multiple evidence combination. In *Proc. SIGIR*. 267–276.

[58] O. Levi, F. Raiber, O. Kurland, and I. Guy. 2016. Selective Cluster-Based Document Retrieval. In *Proc. CIKM*. 1473–1482.

[59] J. Li, C. Huang, X. Wang, and S Wu. 2015. Balancing efficiency and effectiveness for fusion-based search engines in the 'big data' environment. Information Research, 21(2), paper 710. (2015).

[60] S. Liang and M. de Rijke. 2015. Burst-aware data fusion for microblog search. *Inf. Proc. & Man.* 51, 2 (2015), 89–113.

[61] S. Liang, M. de Rijke, and M. Tsagkias. 2013. Late Data Fusion for Microblog Search. In *Proc. ECIR*. 743–746.

[62] S. Liang, I. Markov, Z. Ren, and M. de Rijke. 2018. Manifold Learning for Rank Aggregation. In *Proc. WWW*. 1735–1744.

[63] S. Liang, Z. Ren, and M. de Rijke. 2014. Fusion helps diversification. In *Proc. SIGIR*. 303–312.

[64] S. Liang, Z. Ren, and M. de Rijke. 2014. The Impact of Semantic Document Expansion on Cluster-Based Fusion for Microblog Search. In *Proc. ECIR*. 493–499.

[65] D. Lillis, F. Toolan, R. W. Collier, and J. Dunnion. 2006. ProbFuse: a probabilistic approach to data fusion. In *Proc. SIGIR*. 139–146.

[66] D. Lillis, F. Toolan, R. W. Collier, and J. Dunnion. 2008. Extending Probabilistic Data Fusion Using Sliding Windows. In *Proc. ECIR*. 358–369.

[67] D. Lillis, L. Zhang, F. Toolan, R. W. Collier, D. Leonard, and J. Dunnion. 2010. Estimating Probabilities for Effective Data Fusion. In *Proc. SIGIR*. 347–354.

[68] D. E. Losada, J. Parapar, and A. Barreiro. 2017. Multi-armed bandits for adjudicating documents in pooling-based evaluation of information retrieval systems. *Inf. Proc. & Man.* 53, 5 (2017), 1005–1025.

[69] X. Lu, A. Moffat, and J. S. Culpepper. 2016. The effect of pooling and evaluation depth on IR metrics. *Inf. Retr.* 19, 4 (2016), 416–445.

[70] X. Lu, A. Moffat, and J. S. Culpepper. 2016. Modeling relevance as a function of retrieval rank. In *Proc. AIRS*. 3–15.

[71] X. Lu, A. Moffat, and J. S. Culpepper. 2017. Can deep effectiveness metrics be evaluated using shallow judgment pools?. In *Proc. SIGIR*. 35–44.

[72] J. Mackenzie, F. M. Choudhury, and J. S. Culpepper. 2015. Efficient location-aware web search.. In *Proc. ADCS*. 4.1–4.8.

[73] R. Manmatha and H. Sever. 2002. A formal approach to score normalization for meta-search. In *Proc. of HLT*. 98–103.

[74] I. Markov, A. Arampatzis, and F. Crestani. 2012. Unsupervised linear score normalization revisited. In *Proc. SIGIR*. 1161–1162.

[75] G. Markovits, A. Shtok, O. Kurland, and D. Carmel. 2012. Predicting query performance for fusion-based retrieval. In *Proc. CIKM*.

[76] D. Metzler and W. B. Croft. 2005. A Markov random field model for term dependencies. In *Proc. SIGIR*. 472–479.

[77] M. Montague and J. A. Aslam. 2002. Condorcet fusion for improved retrieval. In *Proc. CIKM*. 538–548.

[78] M. H. Montague and J. A. Aslam. 2001. Relevance Score Normalization for Metasearch. In *Proc. CIKM*. 427–433.

[79] A. Mourao, F. Martins, and J. Magalhaes. 2013. NovaSearch at TREC 2013 Federated Web Search Track: Experiments with rank fusion.. In *Proc. TREC*.

[80] A. Mourao, F. Martins, and J. Magalhaes. 2014. Inverse square rank fusion for multimodal search. In *Proc. CBMI*. 1–6.

[81] K. B. Ng and P. P. Kantor. 1998. An Investigation of the Preconditions for Effective Data Fusion in Information Retrieval: A Pilot Study. (1998).

[82] D. Parikh and R. Polikar. 2007. An ensemble-based incremental learning approach to data fusion. *IEEE Trans. on Systems, Man, and Cybernetics, Part B (Cybernetics)* 37, 2 (2007), 437–450.

[83] T. Qin, X. Geng, and T.-Y. Liu. 2010. A New Probabilistic Model for Rank Aggregation. In *Proc. NIPS*. 1948–1956.

[84] E. Rabinovich, O. Rom, and O. Kurland. 2014. Utilizing relevance feedback in fusion-based retrieval. In *Proc. SIGIR*. 313–322.

[85] F. Raiber and O. Kurland. 2014. Query-performance prediction: setting the expectations straight. In *Proc. SIGIR*. 13–22.

[86] M. Sanderson. 2010. Test Collection Based Evaluation of Information Retrieval Systems. *Found. Trends in Inf. Ret.* 4, 4 (2010), 247–375.

[87] S. B. Selvadurai. 2007. *Implementing a metasearch framework with content-directed result merging.* Master's thesis. North Carolina State University.

[88] D. Sheldon, M. Shokouhi, M. Szummer, and N. Craswell. 2011. LambdaMerge: Merging the results of query reformulations. In *Proc. WSDM*. 795–804.

[89] M. Shokouhi. 2007. Segmentation of Search Engine Results for Effective Data-Fusion. In *Proc. ECIR*. 185–197.

[90] M. Shokouhi and L. Si. 2011. Federated Search. *Found. Trends in Inf. Ret.* 5, 1 (2011), 1–102.

[91] X. M. Shou and M. Sanderson. 2002. Experiments on data fusion using headline information. In *Proc. SIGIR*. 413–414.

[92] A. Shtok, O. Kurland, and D. Carmel. 2016. Query Performance Prediction Using Reference Lists. *ACM Trans. Inf. Sys.* 34, 4 (2016), 19:1–19:34.

[93] S. Tulyakov, S. Jaeger, V. Govindaraju, and D. S. Doermann. 2008. Review of Classifier Combination Methods. In *Machine Learning in Document Analysis and Recognition*. 361–386.

[94] C. C. Vogt. 2000. How much more is better? Characterising the effects of adding more IR Systems to a combination. In *Proc. RIAO*. 457–475.

[95] C. C. Vogt and G. W. Cottrell. 1998. Predicting the Performance of Linearly Combined IR Systems. In *Proc. SIGIR*. 190–196.

[96] C. C. Vogt and G. W. Cottrell. 1999. Fusion via linear combination of scores. *Inf. Retr.* 1, 3 (1999), 151–173.

[97] E. M. Voorhees and D. K. Harman. 2005. *TREC: Experiment and Evaluation in Information Retrieval.* The MIT Press.

[98] W. Webber, A. Moffat, and J. Zobel. 2010. The Effect of Pooling and Evaluation Depth on Metric Stability. In *Proc. EVIA*. 7–15.

[99] S. Wu. 2007. A Geometric probabilistic framework for data fusion in information retrieval. In *Proc. FUSION*. 1–8.

[100] S. Wu. 2009. Applying statistical principles to data fusion in information retrieval. *Expert Syst. Appl.* 36, 2 (2009), 2997–3006.

[101] S. Wu. 2012. Applying the data fusion technique to blog opinion retrieval. *Expert Syst. Appl.* 39, 1 (2012), 1346–1353.

[102] S. Wu. 2012. Linear combination of component results in information retrieval. *Data Knowl. Eng.* 71, 1 (2012), 114–126.

[103] S. Wu. 2013. The weighted Condorcet fusion in information retrieval. *Inf. Proc. & Man.* 49, 1 (2013), 108–122.

[104] S. Wu, Y. Bi, and S. I. McClean. 2007. Regression Relevance Models for Data Fusion. In *Proc. DEXA*. 264–268.

[105] S. Wu, Y. Bi, and X. Zeng. 2011. The Linear Combination Data Fusion Method in Information Retrieval. In *Proc. DEXA*. 219–233.

[106] S. Wu, Y. Bi, X. Zeng, and L. Han. 2009. Assigning appropriate weights for the linear combination data fusion method in information retrieval. *Inf. Proc. & Man.* 45, 4 (2009), 413–426.

[107] S. Wu and F. Crestani. 2002. Data fusion with estimated weights. In *Proc. CIKM*. 648–651.

[108] S. Wu, F. Crestani, and Y. Bi. 2006. Evaluating Score Normalization Methods in Data Fusion. In *Proc. AIRS*. 642–648.

[109] S. Wu and C. Huang. 2014. Search result diversification via data fusion. In *Proc. SIGIR*. 827–830.

[110] S. Wu, J. Li, X. Zeng, and Y. Bi. 2014. Adaptive data fusion methods in information retrieval. *JASIST* 65, 10 (2014), 2048–2061.

[111] S. Wu and S. McClean. 2006. Performance prediction of data fusion for information retrieval. *Inf. Proc. & Man.* 42, 4 (2006), 899–915.

[112] J. Xia, C. Xu, and S. Wu. 2016. Differential Evolution-Based Fusion and Its Properties for Web Search. In *Proc. WISA*. 67–70.

[113] X. Xue and W. B. Croft. 2013. Modeling reformulation using query distributions. *ACM Trans. Inf. Sys.* 31, 2 (2013), 6:1–6:34.

[114] M. Yasukawa, J. S. Culpepper, and F. Scholer. 2015. Data fusion for Japanese term and character n-gram search.. In *Proc. ADCS*. 10.1–10.4.

[115] H. P. Young. 1988. Condorcet's theory of voting. *American Political Science Review* 82, 4 (1988), 1231–1244.

[116] C. Zhang and Y. Ma. 2012. *Ensemble machine learning: methods and applications.* Springer.