

## Curriculum Vitae

### James Allan

Department of Computer Science  
140 Governors Drive  
University of Massachusetts  
Amherst, Massachusetts 01003-4610

413/545-3240  
413/545-1789  
allan@cs.umass.edu  
<http://www.cs.umass.edu/~allan>

Date of Birth: June 19, 1961  
Citizenship: United States

### *Principal fields of interest*

Information retrieval, organization, and visualization; Topic/event Detection and Tracking; Evaluation of retrieval; Information access and capture

### *Education*

**Ph.D.** in Computer Science, Cornell University, 1995. Department of Computer Science. Thesis: Automatic Hypertext Construction, under Professor Gerard Salton. Minor: Theatre Arts (directing).

**M.S.** in Computer Science, Cornell University, 1991.

**A.B.** in Mathematics, Grinnell College, 1983.

### *Employment*

**2008-** Professor, Computer Science, University of Massachusetts Amherst

**2003-** Co-Director, Center for Intelligent Information Retrieval, UMass Amherst

**2003-08** Associate Professor, Computer Science, University of Massachusetts Amherst

**1998-2003** Assistant Professor, Computer Science, University of Massachusetts Amherst.

**1996-2003** Assistant Director, Center for Intelligent Information Retrieval, University of Massachusetts at Amherst (Acting Director, 1999-2000).

**1996-98** Research Assistant Professor, Computer Science, University of Massachusetts Amherst.

**1994-96** Senior Post-doctoral Research Associate, Computer Science, University of Massachusetts at Amherst. Working with the Center for Intelligent Information Retrieval, under the direction of Professor Bruce Croft.

**1989-94** Research Assistant, Computer Science, Cornell University. Working with Professor Gerard Salton on Information Retrieval research using the Smart system. Part of the Information Access and Capture research group.



**Publications (in submission)****Publications (journals)**

1. G. Kumaran and J. Allan, "Adapting Information Retrieval Systems to User Queries." *Information Processing and Management*, special issue on Adaptive Information Retrieval, 44(6):1838-1862, 2008.
2. E.M. Kornfield, R. Manmatha, and J. Allan, "Further Explorations in Text Alignment with Handwritten Documents." *International Journal of Document Analysis and Recognition*, Springer Berlin, Heidelberg, 10(1):39-52, June 2007.
3. J. Allan, V. Lavrenko, and M. Connell, "A Month to Topic Detection and Tracking in Hindi." *ACM Transactions on Asian Language Information Processing*, 2(2):85-10, June 2003 (actually published early 2004).
4. A. Leuski and J. Allan, "Interactive information retrieval using clustering and spatial proximity." *Journal of User Modeling and user-Adapted Interaction*, Kluwer Academic Publishers, 14(2-3):259-288, June 2004.
5. A. Arnt, S. Zilberstein, J. Allan, and A-I Mouaddib, "Dynamic composition of information retrieval techniques." *Journal of Intelligent Information Systems*, Kluwer Academic Publishers, 23(1):67-97, 2004.
6. J. Allan, "Robust Techniques for Organizing and Retrieving Spoken Documents." *Eurasip Journal on Applied Signal Processing*. Hindawi publishers, 2003(2):103-114, February 2003.
7. J. Allan, "Knowledge Management and Speech Recognition" in *IEEE Computer*. April 2002, pages 46-47.
8. J. Allan, "Detection as multi-topic tracking." In *Information Retrieval*, Kluwer Academic Publishers, 5(2/3):139-157, 2002.
9. R.W.P. Luk, A.T.S. Chan, T.S. Dillon, H.V. Leong, W.B. Croft, and J. Allan, "A Survey on Searching and Indexing of XML Documents". *Journal of the American Society of Information Science and Technology*, 53(6):415-437, 2002.
10. J. Allan, A. Leuski, R. Swan, and D. Byrd, "Evaluating combinations of ranked lists and visualizations of inter-document similarity." *Information Processing and Management*, 37(3):435-458, 2001.
11. A. Leuski and J. Allan, "Strategy-based interactive cluster visualization for information." *International Journal of Digital Libraries*, 3(2): 170-184, 2000.
12. D. Rus, J. Allan. "Structural queries in electronic corpora". *Multimedia Tools and Applications on Electronic Publishing*, 6(2):153-170, March 1998.
13. J. Allan. "Building Hypertext Using Information Retrieval". *Information Processing and Management*, 33(2):145-159, 1997.
14. G. Salton, J. Allan, and A. Singhal. "Automatic Text Decomposition and Structuring". *Information Processing and Management*, 32(2):127-138, 1996.
15. G. Salton and J. Allan. "Selective Text Utilization and Text Traversal". *International Journal of Human-Computer Studies*, 43:483-497, 1995.
16. C. Buckley, J. Allan, and G. Salton. "Automatic Routing and Retrieval using Smart: TREC-2". *Information Processing and Management*, 31(3):313-324, 1995.



17. G. Salton, J. Allan, C. Buckley, and A. Singhal. "Automatic Analysis, Theme Generation, and Summarization of Machine-Readable Texts". *Science* 264 (3 June 1994), pp. 1421-1426. Also in *Information Retrieval and Hypertext*, Agosti and Smeaton (Kluwer: 1996), pp. 51-72. Also in *Readings in Information Visualization*, S. Card, J. Mackinlay, and B. Schneiderman, eds., San Francisco: Morgan Kaufmann Publishers, 1999, pp. 413-418.
18. G. Salton, C. Buckley, and J. Allan. "Automatic Structuring and Retrieval of Large Text Files". *Communications of the ACM*, February, 1994.
19. G. Salton, J. Allan and C. Buckley. "Automatic Structuring of Text Files". *Electronic Publishing*, 5(1), March 1992, pp. 1-52.

### **Publications (refereed conferences)**

20. X. Yi and J. Allan, "A Comparative Study of Utilizing Topic Models for Information Retrieval." To appear in Proceedings of ECIR 2009.
21. B. Carterette, V. Pavlu, E. Kanoulas, J. Aslam, and J. Allan, "If I had a Million Queries". To appear in Proceedings of ECIR 2009.
22. M. Lease, J. Allan, and W.B. Croft, "RegressionRank: Learning to Meet the Opportunity of Descriptive Queries." To appear in Proceedings of ECIR 2009.
23. L. Friedland and J. Allan, "Joke Retrieval: Recognizing the same joke told differently." Proceedings of CIKM 2008, pp. 883-892.
24. K. Parton, K. McKeown, J. Allan, and E. Henestroza, "Simultaneous Multilingual Search for Translingual Information Retrieval." Proceedings of CIKM 2008, pp. 719-728.
25. K. Lee, W.B. Croft, and J. Allan, "A Cluster-Based Resampling Method for Pseudo-Relevance Feedback." Proceedings of ACM SIGIR 2008, pp. 235-242.
26. B. Carterette, V. Pavlu, E. Kanoulas, J. Aslam, and J. Allan, "Evaluation over Thousands of Queries." Proceedings of ACM SIGIR 2008, pp. 651-658.
27. G. Kumaran and J. Allan, "Effective and Efficient User Interaction for Long Queries." Proceedings of ACM SIGIR 2008, pp. 11-18.
28. A. Feng and J. Allan, "Finding and Linking Incidents in News." *Proceedings of CIKM 2007*, pp. 821-829.
29. M. Smucker, J. Allan, and B. Carterette, "A Comparison of Statistical Significance Tests for Information Retrieval Evaluation." *Proceedings of CIKM 2007*, pp. 623-632.
30. V. Lavrenko, X. Yi, and J. Allan, "Information Retrieval on Empty Fields." Proceedings of the NAACL HLT Conference, 2007, pp. 89-96.
31. G. Kumaran and J. Allan, "A Case for Shorter Queries, and Helping Users Create Them." Proceedings of the NAACL HLT Conference, 2007, pp. 220-227.
32. B. Schiffman, K. McKeown, R. Grishman, and J. Allan, "Question Answering Using Integrated Information Retrieval and Information Extraction." Proceedings of the NAACL HLT Conference, 2007, pp. 532-539.
33. R. Bekkerman, H. Raghavan, J. Allan, and K. Eguchi, "Interactive Clustering of Text Collections According to a User-Specified Criterion." Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), pp. 684-689, 2007.



34. R. Bekkerman, S. Zilberstein, and J. Allan, "Web Page Clustering using Heuristic Search in the Web Graph." *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2280-2285, 2007.
35. B. Carterette, J. Allan, and R. Sitaraman, "Minimal Test Collections for Retrieval Evaluation." *Proceedings of ACM SIGIR*, pp. 268-283, 2006. **Best paper award**
36. M. Smucker and J. Allan, "Find-Similar: Similarity Browsing as a Search Tool." *Proceedings of ACM SIGIR*, pp. 461-468, 2006.
37. B. Carterette and J. Allan, "Incremental Test Collections." *Proceedings of the Conference on Information and Knowledge Management (CIKM)*, pp. 680-687, 2005.
38. J. Allan, B. Carterette, and J. Lewis, "When Will Information Retrieval Be Good Enough?" *Proceedings of SIGIR*, pp. 433-440, 2005.
39. H. Raghavan and J. Allan, "Matching Inconsistently Spelled Names in Automatic Speech Recognizer Output for Information Retrieval." *Proceedings of HLT/EMNLP*, pp. 251-458, 2005.
40. G. Kumaran and J. Allan, "Using Names and Topics for New Event Detection." *Proceedings of HLT/EMNLP*, pp. 121-128, 2005.
41. J. Allan, S. Harding, D. Fisher, A. Bolivar, S. Guzman-Lara, and P. Amstutz, "Taking Topic Detection From Evaluation to Practice." To appear in *Proceedings of the Hawaii International Conference on System Sciences*, January 2005.
42. R. Nallapati, A. Feng, F. Peng, and J. Allan, "Event Threading Within News Topics." *Proceedings of CIKM*, November 2004, pp. 446-453.
43. G. Kumaran and J. Allan, "Text Classification and Named Entities for New Event Detection." *Proceedings of SIGIR*, pp. 297-304, July 2004.
44. C.H. Gooi and J. Allan, "Cross-Document Coreference on a Large Scale Corpus." *Proceedings of NAACL/HLT*, pp. 9-16, 2004.
45. D. Kelly, F. Diaz, N. Belkin, and J. Allan, "A User-centered Approach to Evaluating Topic Models." *Proceedings of the European Conference on Information Retrieval*, April 2004.
46. E.M. Kornfield, R. Manmatha, and J. Allan, "Text Alignment with Handwritten Documents." *International Workshop on Document Image Analysis for Libraries (DIAL 2004)*, January 2004.
47. J. Allan, A. Feng, and A. Bolivar, "Flexible Intrinsic Evaluation of Hierarchical Clustering for TDT." *Proceedings of CIKM*, pp. 263-270, 2003.
48. J. Allan, C. Wade, and A. Bolivar, "Retrieval and novelty detection at the sentence level." *Proceedings of SIGIR*, pp. 314-321, 2003.
49. R. Nallapati and J. Allan, "Capturing term dependencies using a sentence tree based language model." *Proceedings of CIKM*, pp. 383-390, November 2002.
50. A. Leuski and J. Allan, "Improving realism of topic tracking evaluation." *Proceedings of SIGIR*, pp. 89-96, 2002.
51. J. Allan and H. Raghavan, "Using part-of-speech patterns to reduce query ambiguity." *Proceedings of SIGIR*, pp. 307-314, 2002.
52. V. Lavrenko, J. Allan, E. DeGuzman, D. LaFlamme, V. Pollard, and S. Thomas, "Relevance Models for Topic Detection and Tracking." *Proceedings of the Human Language Technology Conference (HLT)*, pp. 104-110, 2002.



53. J. Allan, V. Khandelwal, and R. Gupta, "Time-based Summarization of News", *Proceedings of SIGIR*, pp. 10-18, 2001. **Best paper award.**
54. J. Allan, V. Lavrenko, and H. Jin, "First story detection in TDT is hard", *Proceedings of CIKM*, pp. 374-181, 2000.
55. V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan, "Language models for financial news recommendation," *Proceedings of CIKM*, pp. 389-396, 2000.
56. A. Leuski and J. Allan, "Lighthouse: Showing the Way to Relevant Information," *Proceedings of the IEEE Symposium on Information Visualization*, IEEE Computer Society, pp. 125-130, 2000.
57. R. Swan and J. Allan, "Automatic Generation of Overview Timelines," *Proceedings of SIGIR*, pp. 49-56, 2000.
58. A. Leuski and J. Allan, "Improving Interactive Retrieval by Combining Ranked List and Clustering." *Proceedings of RIAO*, College de France, pp. 665-681, 2000.
59. R. Swan and J. Allan, "Extracting Significant Time Varying Features from Text," *Proceedings CIKM*, ACM Press, pp. 38-45, 1999.
60. R. Papka and J. Allan, "Document Classification using Multiword Features," *Proceedings of CIKM*, 1998.
61. A. Leouski and J. Allan, "Evaluating a Visual Navigation System for a Digital Library," *Proceedings of European Digital Libraries Conference*, pp. 535-554, 1998.
62. J. Allan, R. Papka, and V. Lavrenko, "On-line New Event Detection and Tracking." *Proceedings of SIGIR*, pp. 37-45, 1998.
63. R. Swan and J. Allan, "Aspect Windows, 3-D Visualizations, and Indirect Comparisons of Information Retrieval Systems." *Proceedings of SIGIR*, pp. 173-181, 1998.
64. M. Hirsch and J. Allan, "A Graphic Interface for User Directed Clustering of Retrieved Documents." *Proceedings of the 1997 Spring Congress*, American Medical Informatics Association, p. 95, 1997.
65. J. Allan. "Incremental Relevance Feedback for Information Filtering". *Proceedings of SIGIR*, pp. 270-278, 1996.
66. J. Allan. "Automatic Hypertext Link Typing". *Proceedings of the ACM Conference on Hypertext*, pp. 42-52, 1996.
67. J. Allan. "Relevance Feedback with Too Much Data". *Proceedings of SIGIR*, pp. 337-343, 1995.
68. J. Allan and D. Rus. "Structural Queries in Electronic Corpora". *Proceedings of DAGS*, 1995.
69. G. Salton and J. Allan. "Automatic Text Decomposition and Structuring". *Proceedings of RIAO*, 1994.
70. C. Buckley, G. Salton, and J. Allan. "The Effect of Adding Relevance Information in a Relevance Feedback Environment". *Proceedings of SIGIR*, 1994.
71. G. Salton and J. Allan. "Text Retrieval Using the Vector Processing Model." *Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval (SDAIR)*, 1994.
72. G. Salton and J. Allan. "Selective Text Utilization and Text Traversal". *Proceedings of ACM Hypertext*, pp. 131-144, 1993.



73. J. Allan and G. Salton. "The Identification of Text Relations Using Automatic Hypertext Linking." Workshop on Intelligent Hypertext, in conjunction with the Second International Conference on Information and Knowledge Management, 1993.
74. J. Allan, J. Davis, D. Krafft, D. Rus, and D. Subramanian. "Information Agents for Building Hyperlinks." Workshop on Intelligent Hypertext, in conjunction with the Second International Conference on Information and Knowledge Management, 1993.
75. G. Salton, J. Allan, and C. Buckley. "Approaches to Passage Retrieval in Full Text Information Systems". *Proceedings of SIGIR*, pp. 49-58, 1993.
76. G. Salton, J. Allan, and C. Buckley. "Automatic Determination of Content Relationships in Natural Language Texts". EP-92 Proceedings, C. Vanoirbeek and G. Coray, eds., Cambridge Univ. Press, Cambridge, England 1992, pp. 165-182.

### **Publications (book chapters)**

77. J. Allan, W.B. Croft, and J. Callan, "The University of Massachusetts and a Dozen TRECs." In *TREC: Experiment and Evaluation in Information Retrieval*, E.M. Voorhees and D.K. Harman, eds. MIT Press, chapter 11, pp. 261-286, 2005.
78. J. Allan, "Modeling Topics for Detection and Tracking" in *Pattern Recognition in Speech and Language Processing*, Wu Chou and Biing Hwang Juang, eds., pp. 353-376. CRC Press, 2003.
79. J. Allan, "Perspectives on Information Retrieval and Speech" in *Information Retrieval Techniques for Speech Applications*, A.R. Coden, E.W. Brown, and S. Srinivasan, eds. Springer-Verlag Lecture Notes in Computer Science, volume 2273, pp. 1-10, 2002.
80. J. Allan, "Introduction to Topic Detection and Tracking" in *Topic Detection and Tracking: Event-based Information Organization*, J. Allan, ed., Kluwer Academic Publishers, pp. 1-16, 2002.
81. J. Allan, V. Lavrenko, and R. Swan, "Explorations Within Topic Tracking and Detection" in *Topic Detection and Tracking: Event-based Information Organization*, J. Allan, ed., Kluwer Academic Publishers, pp. 197-224, 2002.
82. R. Papka and J. Allan, "Topic Detection and Tracking: Event Clustering as a Basis for First Story Detection," in *Advances in Information Retrieval*, W. B. Croft, ed., Kluwer Academic Publishers, pp. 97-126, 2000.

### **Publications (as editor of a volume)**

83. K. Jarvelin, J. Allan, P. Bruza, and M. Sanderson, *Proceedings of SIGIR 2004*, ACM Press, 2004.
84. J. Allan, ed., *Topic Detection and Tracking: Event Based Information Retrieval*, Kluwer Academic Press, 2002.
85. J. Allan, ed., *Proceedings of the First International Conference on Human Language Technology Research (HLT 2001)*, Morgan Kauffman, 2001.



**Publications (workshops)**

86. J. Allan, B. Carterette, J.A. Aslam, V. Pavlu, B. Dachev, and E. Kanoulas, "Million Query Track 2007 Overview." Proceedings of TREC. Notebook version, October 2007. Final version online February 2008, in print December 2008, pp. 85-104.
87. M.D. Smucker, J. Allan, and B. Dachev, "UMass Complex Interactive Question Answering (ciQA) 2007: Human Performance as Question Answerers." Proceedings of TREC. (4 pages) Notebook version, October 2007. On-line version, February 2008.
88. X.Yi and J. Allan, "Indri at TREC 2007: Million Query (1MQ) Track." Proceedings of TREC. (4 pages) Notebook version, October 2007. On-line version, February 2008.
89. M. Smucker and J. Allan, "Measuring the Navigability of Document Networks." SIGIR 2007 workshop on Web Information Seeking and Interaction, 2007.
90. G. Kumaran and J. Allan, "UMass at TREC ciQA." Proceedings of the 2006 Text Retrieval Conference (TREC 2006). Available on-line at <http://trec.nist.gov>.
91. G. Kumaran and J. Allan, "Eliciting Information for Adaptive Retrieval." Proceedings of the First International Workshop on Adaptive Information Retrieval (AIR), pp. 18-19, October 2006.
92. A. Feng and J. Allan, "Combining Evidence from Homologous Datasets." Workshop on New Directions in Multilingual Information Access at SIGIR 2006, August 2006.
93. J. Allan, "HARD Track Overview in TREC 2005: High Accuracy Retrieval From Documents." Proceedings of 2005 Text Retrieval Conference (TREC 2005), NIST special publication 500-266, November 2006.
94. F. Diaz and J. Allan, "When Less is More: Relevance Feedback Falls Short and Term Expansion Succeeds at HARD 2005." Proceedings of the 2005 Text Retrieval Conference (TREC 2005), 2006. Available on-line at <http://trec.nist.gov>.
95. N. Abdul-Jaleel, J. Allan, W.B. Croft, F. Diaz, L. Larkey, X. Li, M.D. Smucker, and C. Wade, "UMass at TREC 2004: Novelty and HARD." *Proceedings of 2004 Text Retrieval Conference (TREC 2004)*, 2005.
96. J. Allan, "HARD Track Overview in TREC 2004: High Accuracy Retrieval from Documents." *Proceedings of TREC 2004*, NIST special publication 500-261, pp. 25-35, 2005.
97. H. Raghavan, J. Allan, and A. McCallum, "An Exploration of Entity Models, Collective Classification and Relation Description." *Proceedings of the Second International Workshop on Lnk Analysis and Group Detection (LinkKDD)*, at ACM SIGKDD, pp. 1-10, August 2004.
98. H. Raghavan and J. Allan, "Using Soundex Codes for Indexing ASR Documents." *Proceedings of the Workshop on Interdisciplinary Approaches to Speech Indexing and Retrieval*, at HLT-NAACL, pp. 22-26, 2004.
99. J. Allan, "HARD Track Overview in TREC 2003: High Accuracy Retrieval from Documents." *Proceedings of TREC 2003*, NIST special publication 500-255, pp. 24-37, 2004.
100. N. Abdul-Jaleel, A. Corrada-Emmanuel, Q. Li, X. Liu, C. Wade, and J. Allan, "UMass at TREC 2003: HARD and QA." *Proceedings of TREC 2003*, NIST special publication 500-255, pp. 715-725, 2004.



101. R. Nallapati and J. Allan, "An Adaptive Local Dependency Language Model: Relaxing the Naïve Bayes' Assumption." *Workshop on Mathematical and Formal Methods in Information Retrieval* at SIGIR 2003.
102. L. Larkey, J. Allan, M. Connell, A. Bolivar, and C. Wade, "UMass and TREC 2002: Cross Language and Novelty Tracks" in *Proceedings of TREC 2002*, pp. 43-55.
103. J. Allan, V. Lavrenko, and R. Nallapti, "UMass at TDT 2002" in *Proceedings of the Topic Detection and Tracking Workshop*, 2002.
104. J. Allan, R. Gupta, and V. Khandelal, "Topic Models for Summarizing Novelty" in *Proceedings of the Workshop on Language Modeling in Information Retrieval*, Carnegie Mellon University, pp. 66-71, 2001.
105. J. Allan, M. Connell, W.B. Croft, F. Feng, D. Fisher, and X. Li, "INQUERY and TREC-9", *Proceedings of TREC-9*, NIST special publication, pp. 551-577, 2001.
106. J. Allan, V. Lavrenko, D. Frey, and V. Khandelwal, "UMass at TDT 2000", *Proceedings of the Topic Detection and Tracking Workshop*, 2000.
107. V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan, "Mining concurrent text and time series," *Proceedings of KDD conference workshop on text mining*, pp. 37-44, 2000.
108. J. Allan, V. Lavrenko, D. Malin, and R. Swan, "Detection, Bounds, and Timelines: UMass and TDT- 3." *Proceedings of Topic Detection and Tracking (TDT 1999)*, pp. 167-174, 2000.
109. J. Allan, J. Callan, F. Feng, and D. Malin, "INQUERY and TREC-8." *Proceedings of TREC-8 (1999)*, NIST special publication, pp. 637-644, 2000.
110. R. Papka, J. Allan, and V. Lavrenko, "UMass Approaches to Detection and Tracking at TDT-2." *Proceedings of the Broadcast News Transcription and Understanding Workshop*, San Francisco: Morgan Kaufmann (DARPA), pp. 111-116, 1999.
111. J. Allan, J. Callan, M. Sanderson, J. Xu, and S. Wegmann, "INQUERY at TREC-7." *Proceedings of TREC-7 (1998)*, NIST special publication, pp. 201-216, 1999.
112. J. Allan, J. Carbonell, G. Doddington, J. Yamron, and Y. Yang, "Topic Detection and Tracking Pilot Study: Final Report." *Proceedings of the Broadcast News Transcription and Understanding Workshop*, pp. 194-218, 1998.
113. J. Allan, J. Callan, W.B. Croft, L. Ballesteros, D. Byrd, R. Swan, and J. Xu, "INQUERY Does Battle with TREC-6," *Proceedings of TREC-6 (1997)*, NIST special publication, pp. 169-206, 1998.
114. J. Allan, J. Callan, B. Croft, L. Ballesteros, J. Broglio, J. Xu, and H. Shu. "INQUERY at TREC-5". *Proceedings of TREC-5 (1996)*, NIST special publication, pp. 119-132, 1998.
115. J. Allan, L. Ballesteros, J. Callan, W.B. Croft, and Z. Lu. "Recent Experiments with INQUERY". *Proceedings of TREC-4 (1995)*, NIST special publication, pp. 49-64, 1996.
116. J. Allan and D. Rus. "Does Navigation Require More Than One Compass?". AAI 1995 Fall Symposium workshop on "AI Applications in Knowledge Navigation and Retrieval," 1995.





117. C. Buckley, G. Salton, J. Allan, and A. Singhal. "Automatic Query Expansion Using SMART: TREC-3". The Third Text REtrieval Conference (TREC-3); November, 1994; published April, 1995, pp. 69-80.
118. C. Buckley, J. Allan, and G. Salton. "Automatic Routing and Ad-hoc Retrieval Using SMART: TREC-2". The Second Text REtrieval Conference (TREC-2), August 1993; published March 1994, pp. 45-55.
119. C. Buckley, G. Salton, and J. Allan. "Automatic Retrieval With Locality Information Using SMART". The First Text REtrieval Conference (TREC-1), November 1992, pp. 59-72; published March, 1993.

### **Publications (refereed posters)**

120. E. Aktolga, M.A. Cartright, and J. Allan, "Cross-document Cross-lingual Coreference Retrieval." Poster published in Proceedings of CIKM 2008, pp. 1359-1360.
121. X. Yi and J. Allan, "Evaluating topic models for information retrieval." Poster published in Proceedings of CIKM 2008, pp. 1431-1432.
122. G. Kumaran and J. Allan, "Selective User Interaction." Poster published in the *Proceedings of CIKM*, pp. 923-926. 2007.
123. B. Carterette and J. Allan, "Semiautomatic Evaluation of Retrieval Systems Using Document Similarities." Poster published in the *Proceedings of CIKM*, pp. 873-876. 2007.
124. N. Balasubramanian, J. Allan, and W.B. Croft, "A Comparison of Sentence Retrieval Techniques." Poster published in Proceedings of the ACM SIGIR, 2007, pp. 813-4.
125. X. Yi, J. Allan, and W.B. Croft, "Matching Resumes and Jobs Based on Relevance Models." Poster published in Proceedings of the ACM SIGIR, 2007, pp. 809-10.
126. M.D. Smucker and J. Allan, "Using Similarity Links as Shortcuts to Relevant Web Pages." Poster published in the Proceedings of ACM SIGIR 2007, p. 863-4.
127. G. Kumaran and J. Allan, "Information Retrieval Techniques for Templated Queries." Poster published in the Proceedings of RIAO 2007, p. 104.
128. B. Carterette and J. Allan, "Research Methodologies in Studies of Information Retrieval Evaluation." Poster published in the Proceedings of RIAO 2007, p. 112.
129. X. Yi, J. Allan, and V. Lavrenko, "Discovering Missing Values in Semi-Structured Databases." Poster published in the Proceedings of RIAO 2007, poster number 113, pages 1-15.
130. M. Smucker and J. Allan, "Lightening the Load of Document Smoothing for Better Language Modeling Retrieval." Poster published in the Proceedings of ACM SIGIR 2006, pp. 699-700, 2006.
131. G. Kumaran and J. Allan, "Simple Questions to Improve Pseudo-Relevance Feedback Results." Poster published in the Proceedings of ACM SIGIR 2006, pp. 661-662, 2006.
132. R. Nallapati, W.B. Croft, and J. Allan, "Relevant Query Feedback in Statistical Language Modeling." Poster in *Proceedings of CIKM*, pp.560-563, 2003.
133. J. Allan and G. Kumaran, "Stemming in the language modeling framework." Poster in *Proceedings of SIGIR*, pp. 455-456, 2003.



134. R. Manmatha, A. Feng, and J. Allan, "A critical examination of TDT's cost function." Poster in *Proceedings of SIGIR*, pp. 403-404, 2002.
135. V. Khandelwal, R. Gupta, and J. Allan, "An Evaluation Scheme for Summarizing Topic Shifts in News Streams", poster appearing in *Proceedings of the Human Language Technology Conference (HLT)*, 2001.
136. A. Leouski and J. Allan, "Visual Interactions with a Multidimensional Ranked List." Poster presentation in *Proceedings of SIGIR*, 1998.

### **Publications (not refereed)**

137. G. Kumaran and J. Allan, "Simple Questions for Interactive Information Retrieval." CIIR Technical report IR-507, 2006.
138. Smucker, M.D., Kulp D., and Allan J., "Dirichlet Mixtures for Query Estimation in Information Retrieval," University of Massachusetts, Amherst, CIIR Technical Report IR-445, 2005.
139. G. Kumaran, J. Allan, and A. McCallum, "Classification Models for New Event Detection." CIIR technical report IR-362, 2004.
140. H. Raghavan and J. Allan, "Proper Names and Their Spelling Variations." CIIR technical report IR-361, 2004.
141. R. Nallapati, J. Allan, and S. Mahadavan, "Extraction of Key-Words from News Stories." CIIR technical report IR-345, 2004.
142. H. Raghavan and J. Allan, "Evaluating Entity Models on the TREC Question Answering Task." CIIR technical report IR-342, 2004.
143. B. Carterette, J. Allan, and W.B. Croft, "Ranking Document Retrieval Systems with BLEU." CIIR technical report IR-327, 2003.
144. H. Raghavan, A. McCallum, and J. Allan, "Using Markov Random Fields for Relational Classification in Social Networks Mined from Text." CIIR technical report IR-300, 2003.
145. J. Allan and H. Raghavan, "A Probabilistic Model of Named Entities: You Are What They Say You Are." CIIR technical report IR-296, 2003.
146. J. Allan (editor), J. Aslam, N. Belkin, C. Buckley, J. Callan, B. Croft (editor), S. Dumais, N. Fuhr, D. Harman, D.J. Harper, D. Hiemstra, T. Hofmann, E. Hovy, W. Kraaij, J. Lafferty, V. Lavrenko, D. Lewis, L. Liddy, R. Manmatha, A. McCallum, J. Ponte, J. Prager, D. Radev, P. Resnik, S. Robertson, R. Rosenfeld, S. Roukos, M. Sanderson, R. Schwartz, A. Singhal, A. Smeaton, H. Turtle, E. Voorhees, R. Weischedel, J. Xu, and C. Zhai, "Challenges in Information Retrieval Language Modeling." *SIGIR Forum*, 37:1, spring 2003.
147. J. Allan and G. Kumaran, "Details on Stemming in the Language Modeling Framework." CIIR technical report IR-284, 2003.
148. R. Nallapati and J. Allan, "Sentence-Forest Language Model: A Graph-theoretic Framework to Capture Local Term Dependencies." CIIR technical report IR-281, 2003.
149. F. Diaz and J. Allan, "Browsing-based User Language Models for Information Retrieval." CIIR technical report IR-279, 2003.
150. J. Allan and H. Raghavan, "Entity Models: Construction and Applications." CIIR technical report IR-275, 2003.



151. J. Allan, A. Bolivar, and C. Wade, "Using text categorization to measure the impact of news on public opinion." CIIR technical report IR-264, 2002.
152. J. Allan, V. Lavrenko, and H. Jin, "Comparing effectiveness in TDT and IR." CIIR Technical Report, 2000.
153. J. Allan, H. Jin, M. Rajman, C. Wayne, D. Gildea, V. Lavrenko, R. Hoberman, and D. Caputo, "Topic-based Novelty Detection: 1999 Summer workshop at CLSP, Final Report", available on-line at <http://www.clsp.jhu.edu>.
154. A. Leuski and J. Allan, "The best of both worlds: Combining ranked list and clustering", CIIR technical report, 1999.
155. J. Allan, V. Lavrenko, and P. Mularski, "Measuring query-focused web crawlers", CIIR technical report, 1999.
156. R. Swan, J. Allan, and D. Byrd, "Evaluating a visual information retrieval interface: AspInquery at TREC-6." Position paper for CHI 1998 workshop on information exploration. CIIR technical report, 1998.
157. R. Swan, J. Allan, and D. Byrd, "Evaluating a task-specific information retrieval interface", CIIR technical report, 1998.
158. J. Allan, V. Lavrenko, and R. Papka, "Event Tracking," CIIR technical report, 1998.
159. R. Papka and J. Allan, "On-Line New Event Detection using Single Pass Clustering." UMass Computer Science technical report 98-21 and CIIR technical report, 1998.
160. D. Byrd, R. Swan, and J. Allan, "TREC-6 Interactive Track Report Part I: Experimental Procedure and Initial Results," CIIR technical report, 1997.
161. J. Allan, A. Leouski, R. Swan, "Interactive Cluster Visualization for Information Retrieval." CIIR technical report, 1997.
162. R. Papka and J. Allan, "Why Bigger Windows Are Better Than Smaller Ones." CIIR Technical Report, 1997.
163. R. Swan and J. Allan, "Improving Interactive Information Retrieval Effectiveness with 3-D Graphics," CIIR Technical Report, 1996.
164. J. Allan. "Automatic Hypertext Construction". Ph.D. Thesis. January, 1995.
165. G. Salton, C. Buckley, and J. Allan. "Selective Use of Full-Text Databases". Cornell Computer Science Technical Report 92-1300.

### **Other publications**

166. D. Frey, R. Gupta, V. Khandelwal, V. Lavrenko, A. Leuski, and J. Allan, "Monitoring the News: a TDT demonstration system", demonstration appearing in *Proceedings of the Human Language Technology Conference (HLT)*, pp. 351-355, 2001.
167. R. Swan and J. Allan, "TimeMine: Visualizing automatically constructed timelines," *Proceedings of SIGIR*, ACM Press, 2000, p. 393. Description of a demonstration system presented at the conference.
168. J. Allan "Readings in Information Retrieval by K. Sparck Jones and P. Willett (eds)." Book review published in *Information Processing and Management*, 34(4):4, 1998, pp. 489-490.
169. J. Allan and A. Singhal. "Document Handling." Article in *Encyclopedia of Electrical and Electronics Engineering*. Forthcoming, 1998.



170. M. Agosti and J. Allan. "Introduction to the special issue on methods and tools for the automatic construction of hypertext." *Information Processing and Management*, 33:2, pp. 129-131, 1997.



## Grants

- J. Allan (PI), W.B. Croft (co-PI). Learning Word Relationships Using TupleFlow, NSF CLUE program. \$500,000 total (\$331,434 direct; \$168,566 indirect).
- R. Manmatha (PI), J. Allan (co-PI), E. Berger (co-PI), D. Kulp (co-PI). Cluster Acquisition for Computational Research into Large Scale Data Rich Problems, NSF CISE MRI. \$350,000 total (all direct).
- J. Allan (PI), W.B. Croft (Co-PI), A. McCallum (co-PI), Supporting Distillation for GALE, SRI International (from DARPA), 9/29/2005 to 12/31/2008, \$5,932,714 total (\$3,912,917 direct; \$2,019,797 indirect)
- J. Allan (PI), R. Manmatha (co-PI), A. McCallum (co-PI), Shared Computing Resource for Language and Multimedia Technologies, AFOSR, 05/01/05 to 04/30/06, \$117,321 total (\$117,321 direct; \$0 indirect)
- J. Allan (PI), W.B. Croft (co-PI), Collaborative Project: Core Integration – Leading NSDL toward Long-Term Success, NSF (subcontract to UCAR), 10/1/2002 to 9/30/2006, \$1,717,745 total (\$1,128,294 direct; \$589,451 indirect)
- W.B. Croft (PI), J. Allan (co-PI), Formal Frameworks and Empirical Evaluations for Information Organization, DARPA/SPAWARSYSCEN-SD, 4/10/2002 to 4/9/2006, \$4,353,018 total (\$2,855,980; \$1,497,038 indirect)
- J. Allan (PI), W.B. Croft (co-PI), Search and Browsing Support for NSDL, NSF (subcontract to UCAR), 10/1/2001 to 9/30/2003, \$437,500 total (\$288,502 direct; \$148,998 indirect)
- W.B. Croft (PI), J. Allan (co-PI), Relevance Models and Answer Granularity for Question Answering, DoD, 9/29/2001 to 9/28/2003. \$749,878 total (\$550,772 direct; \$199,106 indirect)
- J. Allan (PI), Links between News Reporting and Opinion Polls, Historical Research Foundation, 9/5/2001 to 7/1/2002, \$19,865 total (\$16,554 direct; \$3,311 indirect)
- W.B. Croft (PI), J. Allan (co-PI), Supporting Effective Access through User- and Topic-based Language Models, NSF, 10/1/2000 to 9/30/2004. \$351,964 total (\$227,073 direct; \$124,891 indirect)
- D. Jensen (PI), J. Allan (co-PI), The Application of Information Mining Technology to the Total Army Injury and Health Outcomes Database, 9/1/2000 to 2/28/2001. \$75,000 total (\$48,860 direct; \$26,140 indirect)
- J. Allan (PI), SNOMED Coding System, 9/1/2000 to 2/28/2001. \$20,910 total (\$13,622 direct; \$7,288 indirect)
- J. Allan (PI), Getting What You Want: Accurate Document Filtering in a W.B. Croft (PI), J. Allan (co-PI) A Unified Framework for Information Retrieval and Management Based on Statistical Language Models, Carnegie Mellon University (prime contractor to DoD), 6/30/2000 to 6/29/2001. \$405,958 total (\$264,468 direct; \$141,490 indirect)
- J. Allan (PI), W.B. Croft (co-PI), D. Jensen (co-PI), Novelty Detection and the Integration of Speech and Text within the Topic Detection and Tracking



- Initiative, Dragon Systems (DARPA subcontract), 6/26/2000 to 12/31/2003. \$150,000 total (\$96,744 direct; \$53,206 indirect)
- J. Allan (PI), Scrollbar Visualization Research Project, 1/10/2000 to 2/9/2000. \$2,250 total (\$1,452 direct; \$798 indirect)
  - S. Zilberstein (PI), J. Allan (co-PI), Dynamic Composition of Information Retrieval Techniques, NSF, 1/1/2000 to 12/31/2002. \$437,184 total (\$284,715 direct; \$152,469 indirect)
  - W.B. Croft (PI), J. Allan (co-PI), Tools for Rapidly Adaptable Translingual Information Retrieval and Organization, DARPA/SPAWARSYSCEN-SD, 7/1/1999 to 6/30/2003. \$1,433,514 total (\$947,058 direct; \$486,456 indirect)
  - J. Allan (PI), W.B. Croft (co-PI), R. Manmatha (co-PI), Involving Undergraduates in Research at the Center for Intelligent Information Retrieval, NSF, 3/1/1999 to 1/28/2002. \$193,728 total (\$167,088 direct; \$26,640 indirect)
  - Manmatha (PI), J. Allan (co-PI), Infrastructure for Large-Scale Multimedia Information Indexing, Retrieval, and Organization, Air Force Office of Scientific Research, 3/1/1999 to 2/29/2000 \$247,600 total (\$247,600 direct; \$0 indirect)
  - Terabyte World, Air Force Research Laboratories, 6/24/1998 to 12/31/2001. \$275,690 total (\$242,020 direct; \$33,670 indirect) This work is subcontracted to Carnegie Mellon University.
  - W.B. Croft (PI) J. Allan(co-PI), Topic Tracking in Broadcast Data Using Hidden Markov Models, Dragon Systems (prime contractor to DARPA), 6/9/1997 to 12/31/1999. \$288,147 total (\$188,331 direct; \$99,816 indirect)
  - J. Allan (PI), R. Manmatha (co-PI), Multimodal Indexing, Retrieval, and Browsing, 3/1/97 to 2/28/02. \$769,111 total (\$527,826 direct; \$241,275 indirect)

### **Recent professional activities**

#### Editorships:

- Associate Editor, ACM Transactions on Information Systems, 2002-
- Associate Editor, Information Processing and Management (Elsevier), 2008-

#### Conference or workshop organization:

- SIGIR
  - 2009, general co-chair
  - 2004, program co-chair
  - 2002, poster chair
  - 2000, panel and demonstration chair
  - 1997, co-chaired Workshop on Info. Reduction (with A. Singhal)
  - 1995, co-chaired Workshop on Automatic Hypertext Construction (with M. Agosti)
- ACL
  - 2007, program co-chair for IR
  - 1999, co-chair of thematic session
- TREC
  - Program Committee, 1999 to 2008
  - 2007-2008, coordinator of Million Query Track



- 2005-2006, coordinator of High Accuracy Retrieval track
- NAACL HLT 2007, workshops and tutorials co-chair
- HLT (Human Language Technologies): 2001, general and program chair
- TDT 2003, 2002, 2001, 2000, 1999, workshop chair all years

Program committees:

- SIGIR senior program committee: 2007, 2006, 2005, 2003, 2001, 1998
- SIGIR program committee: 2008, 2002, 2000, 1999
- CIKM 1998 to 2007 (excepting 2005)
- ECIR 2009
- WSDM 2009, WSDM late breaking results 2009
- NAACL/HLT 2006
- NTCIR 2007
- InfoScale 2007
- ECIR 2007
- HICSS 2006 (mini-track on IR and Digital Library Applications)
- Document Recognition and Retrieval (DRR) 2005
- Human Language Technologies (HLT) Conference 2002
- Broadcast News Transcription and Understanding Workshop, 1999
- KDD 2000 (industrial track)
- ICML 2000
- AAI 2000
- FQAS 2000
- SPIRE 2007, 2000
- ICTAI 1999

**Awards and honors**

- Best paper award, SIGIR 2006.
- Best paper award, SIGIR 2001.
- Nominated for Distinguished Teaching Award, 2000
- Lilly Teaching Fellow, 1999-2000

**Teaching experience**

Spring 2009, CMPSCI 791MT, Mining text and images in digital libraries using grid computing (with R. Manmatha and D. Smith)  
Fall 2008, CMPSCI 187, Data structures in Java  
Spring 2008, CMPSCI 791U, Advanced Topics in Information Retrieval (seminar)  
Fall 2007, CMPSCI 646, Information Retrieval (PhD-level)  
Spring 2007, CMPSCI 187  
Fall 2006, CMPSCI 646  
Spring 2005, CMPSCI 646 and CMPSCI 546, Web Search Engines  
Fall 2004, CMPSCI 646  
Fall 2003, CMPSCI 646  
Spring 2003, CMPSCI 791, Advanced topics in IR



Fall 2002, CMPSCI 646 and CMPSCI 791, Advanced topics in IR (with W.B. Croft)  
Spring 2002, CMPSCI 546, Web Search Engines  
Fall 2000, CMPSCI 445, Information Systems (databases)  
Spring 2000, CMPSCI 646  
Fall 1999, CMPSCI 445 (plus an honors section on database construction) and CMPSCI 791M, Information Mining (with D. Jensen)  
Fall 1998, CMPSCI 445 (plus an honors section on database construction)  
Spring 1997, CMPSCI 646 (co-taught with J. Callan)  
Summer 1993, CS 414, Operating System Principles (Cornell University)  
Summer 1991, CS 414 (Cornell University)

### **Department and University service**

- Graduate Program Director, 2006-
- Graduate Program Committee, 2004-5
- Space Committee (chair), 2004-5
- Computing Committee (chair), 2002-2004
- Special events committee, 1998-2003 (chair, 2000-2003)
- Personnel committee, 1999-2000
- Admissions committee, 2000-2001
- Executive committee, 2000-2001, *ex officio* 2006-
- Ad-hoc campus-wide committee on Bioinformatics and Genomics
- Panel presentation for Welcoming program for new faculty run by the Center for Teaching, 2000.

### **Students supervised**

- Completed, PhD, chair or co-chair of committee
  - Ben Carterette, 2008
  - Mark Smucker, 2008
  - Ao Feng, 2008
  - Giridhar Kumaran, 2008
  - Fernando Diaz, 2007
  - Ron Bekkerman, 2007
  - Hema Raghavan, 2007
  - Ramesh Nallapati, 2006
  - Victor Lavrenko, co-chair (with W.B. Croft), 2004
  - Anton Leouski, chair, 2001.
  - Russell Swan, chair. Died unexpectedly, 2000.
  - Ron Papka, co-chair (with W.B. Croft), 1999.
- Completed, PhD, member of committee
  - Trevor Strohman, 2008
  - Shaolei Feng, 2008
  - Xiaoyong Liu, 2008
  - Yun Zhou, 2008





- Jiwoon Jeon, 2007
- Don Metzler, 2007
- Vanessa Murdock, 2006
- Xiayan Li, 2006
- Tom Kalt, 2005
- Jeremy Pickens, 2004
- Dawn Lawrie, 2003
- Jay Ponte, 1999.
  
- Completed, terminal MS
  - Courtney Wade, 2005
  - Alvaro Bolivar, 2004
  - Micah Kornfield, 2004
  - Chooi Chung Geong, 2004
  - Rahul Gupta, 2001.
  - Vikash Khandelwal, 2001.
  - Nitin Srimal, 1999.
  - Sudheer Gaddam, 2000.
  
- Current
  - Niranjan Balasubramanian, research advisor
  - Xing Yi, research advisor
  - Elif Aktolga, research advisor
  - Henry Field, research advisor
  - Marc Cartright, research advisor
  - Jeff Dalton, research advisor
  
  - Evangelos Kanoulas, Northeastern University
  - Matt Lease, Brown University
  
- Non-UMass, completed
  - Soyeon Park, outside member of Ph.D. committee (Rutgers University). Completed January 1999.
  - Katya Pelekhov, outside member of Ph.D. committee (Dartmouth College). Completed 2000.
  - Mark Montague, outside member of Ph.D. committee (Dartmouth College). Completed 2002.
  
- Undergraduate research and honors projects
  - Blagovest Dachev (honors project, 2007-8); Jennifer Neville (honors thesis committee, 2000); Paul Ogilvie (chair of honors thesis committee, 2000; research, 1998-9); Peter Mularski (chair of honors thesis committee, 1999); Lida Tang (honors thesis, 1999); Matthew Billmers (honors course, 1998); Lida Tang (honors course, 1998)

