Shafiq Rayhan Joty

Curriculum Vitae
April 8, 2013

Department of Computer Science
University of British Columbia
2366 Main Mall, Vancouver, B.C.
Canada V6T 1Z4

Phone: (778) 320-5689
Email: rjoty@cs.ubc.ca
Web:www.cs.ubc.ca/~rjoty

Research Interest

- **Natural Language Processing / Computational Linguistics**

    - **Computational Discourse** (Topic Modeling, Dialog Act Modeling, Discourse Parsing)

    - **Text Summarization and Question Answering**

    - **Machine Learning** (Graphical Models and Its application to NLP problems)

Education

- **Ph.D. in Computer Science, expected Mar, 2013**   **Sep 2008 - Present**
  University of British Columbia
  Vancouver, B.C., Canada
  Thesis title: Discourse Analysis of Asynchronous Conversations (e.g., blog, email)
  Advisor: Dr. Giuseppe Carenini
  Committee members: Prof. Raymond T. Ng and Prof. Nando de Freitas
- **M.Sc. in Computer Science**   **Sep 2006 - Apr 2008**
  University of Lethbridge
  Thesis title: Answer Extraction for Simple and Complex Questions
  Advisor: Prof. Yllias Chali
- **B.Sc. in Computer Science and Information Technology**   **Jan 2001 - Aug 2005**
  CGPA 4.97/5.00, **Rank 2nd** in the merit list
  Islamic University of Technology (IUT), Gazipur, Bangladesh.
  Thesis title: Design and Development of a Communication Server for IUT
  Advisor: Prof. M. A. Mottalib, Head, CIT Department.

Research Experience

- **Research Assistant,** Computer Science Dept., Univ. of British Columbia   *Sep 2008 - Present*
- ***Research Intern,*** *Microsoft Research Asia*   *Sep 2010 – Dec 2010*
    Web Intelligence Group, *Mentor:* Dr. Chin-Yew Lin
- **Research Assistant,** Computer Science Dept., Univ. of Lethbridge   *Sep 2006 – Aug 2008*

Teaching Experience

- **Teaching Assistant,** Computer Science Dept., Univ. of British Columbia   *Sep 2008 – Present*
  Courses TAed: Grad AI, Ugrad AI, Object Oriented Programming, Intro. to Comp.
- **Teaching Assistant,** Computer Science Dept., Univ. of Lethbridge   *Sep 2006 – Aug 2008*
  Courses TAed: Introduction to Computers
- **Lecturer,** Computer Science Dept., Islamic Univ. of Technology   *Dec 2005 – Aug 2006*
  Courses Taught: Theory of computation, UNIX prog., Systems prog., OOP

Other Related Experience

- ***Software Developer (Part time),*** Anam Technologies Inc, Bangladesh   May 2002 – Dec 2005

Impact and Visibility

- **BC3 Email and Blog Corpora annotated with topics:**
  Will be publicly available soon at www.cs.ubc.ca/nest/lci/bc3.html
- **Topic Segmenter & Labeler for Asynchronous Conversations:**
  Currently being used at Microsoft Research Asia.
- **Discourse Parsers (Sentence-level and Document-level):**
  Will be made available only for research purposes at http://www.cs.ubc.ca/~rjoty/

Major Honors and Awards

- Microsoft Research Excellent Intern
  Microsoft Research, Dec 2010
- NSERC Alexander Graham Bell Canada Graduate Scholarship (CGS D)
  May 2010 - Apr 2012 [$35,000 + $5000 (top up) per year]
- Four Year Fellowship (FYF) at UBC
  Sep 2008 – Aug 2012 [$18, 000 per year]
- Faculty of Science Graduate Award at UBC
  Sep 2008 – Aug 2012 [$7, 200 per year]
- Governor General's Gold Medal nominee
  University of Lethbridge, May 2008
- David R. Cheriton Scholarship, University of Waterloo, 2008 (Declined)
- FS Chia Ph.D. Scholarship at the University of Alberta, 2008 (Declined)

Invited Talks

- NSERC Business Intelligence Network AGM, IBM Software Labs, Canada, (Nov, 2012)
  Title: Intra- and Multi-sentential Rhetorical Parsing for Document-level Discourse Analysis
- NW-NLP workshop, Microsoft Research Redmond (May, 2012)
  Title: A Novel Discriminative Framework for Sentence-Level Discourse Analysis
- Business Intelligence (BI) Workshop-NSERC BIN Watson, IBM Watson center, (Feb, 2012)
  Title: Discourse Analysis of Asynchronous Conversations (e.g., emails, blogs)
- UBC Social Network Group (Feb, 2012)
  Title: Discourse Analysis of Asynchronous Conversations (e.g., emails, blogs)
- UBC Laboratory for Computational Intelligence (LCI) Forum Talk (Apr, 2011)
  Title: Unsupervised Modeling of Dialog Acts in Asynchronous Conversations
- Microsoft Research Asia (Nov, 2010)
  Title: Topic and Dialog Act Modeling of Asynchronous Conversations
- UBC Laboratory for Computational Intelligence Forum (LCI) Talk (Feb, 2010)
  Title: "What are they talking about?" Finding Topics in Email Conversations
- University of Alberta NLP Group (Jan, 2008)
  Title: Answering Simple and Complex Questions
- University of Lethbridge colloquium (Nov, 2007)

Professional Activities

- Reviewer: IJCAI-2013, IUI-2011, NAACL-2012, ACL-2012, EMNLP-2012.
- Associate VP position in the Bangladesh Students' Association, UBC (2009-2010).
- Organizing committee member in the ICCIT 2005 conference, held at IUT.

Professional Membership

- Association for Computational Linguistics (ACL)
- ACL special interest group for linguistic data and corpus-based approaches to NLP (SIGDAT)
- Association for the Advancement of Artificial Intelligence (AAAI)

Publications

Refereed Journal Publications

[1] **Shafiq Joty,** Giuseppe Carenini and Raymond Ng**.** Topic Segmentation and Labeling in Asynchronous Conversations. *Journal of AI Research (**JAIR**) 2013 [To Appear].*

[2] Yllias Chali, **Shafiq Joty** and Sadid Hasan**.** Complex Question Answering: Unsupervised Learning Approaches and Experiments. *Journal of AI Research (**JAIR**)* 2009. Vol. 35, Page 1-47.

[3] Yllias Chali, **Shafiq Joty** and Sadid Hasan. Improving Graph-based Random Walks for Complex Question Answering Using Syntactic and Shallow Semantic Kernels. *Information Processing and Management (**IPM**), Special Issue on Question Answering* 2010. Elsevier.

Refereed Conference Publications

[1] **Shafiq Joty**, Giuseppe Carenini, Raymond Ng and Yashar Mehdad. Combining Intra- and Multi-sentential Rhetorical Parsing for Document-level Discourse Analysis. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (**ACL**)* 2013, Sofia, Bulgaria.

[2] Yashar Mehdad, Giuseppe Carenini, Raymond Ng and **Shafiq Joty**. Towards Topic Labeling with Phrase Entailment and Aggregation. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL-HLT**)* 2013. Atlanta, USA.

[3] **Shafiq Joty**, Giuseppe Carenini and Raymond Ng. A Novel Discriminative Framework for Sentence-Level Discourse Analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the Conference on Natural Language Learning (**EMNLP-CoNLL**)* 2012, Jeju, Korea.

[4] **Shafiq Joty,** Giuseppe Carenini, and Chin-Yew Lin. Unsupervised Modeling of Dialog Acts in Asynchronous Conversations. In *Proceedings of the twenty second International Joint Conference on Artificial Intelligence (**IJCAI**)* 2011. Barcelona, Spain.

[5] **Shafiq Joty,** Giuseppe Carenini, Gabriel Murray and Raymond Ng. Supervised Topic Segmentation of Email Conversations. In *Fifth International AAAI Conference on Weblogs and Social Media (**ICWSM**)* 2011. Barcelona, Spain. (Short paper).

[6] Nicholas FitzGerald, Giuseppe Carenini, Gabriel Murray and **Shafiq Joty**. Exploiting Conversational Features to Detect High-Quality Blog Comments. In *Proceedings of the Canadian Conference on Artificial Intelligence (**CAI**)* 2011. St. Johns, Newfoundland. (Short paper).

[7] **Shafiq Joty,** Giuseppe Carenini, Gabriel Murray and Raymond Ng. Exploiting Conversation Structure in Unsupervised Topic Segmentation for Emails. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP**)* 2010. Cambridge, Massachusetts, USA.

[8] Yllias Chali and Sadid Hasan, **Shafiq Joty**. Do Automatic Annotation Techniques Have Any Impact on Supervised Complex Question Answering? In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing (**ACL-IJCNLP**)* 2009. Suntec, Singapore. (Short paper).

[9] Yllias Chali and Sadid Hasan, **Shafiq Joty**. A SVM-Based Ensemble Approach to Multi-Document Summarization. In *Proceedings of the 22nd Canadian Conference on Artificial Intelligence (**CAI**)* 2009, Kelowna, B.C. Lecture Notes in Artificial Intelligence, Springer-Verlag.

[10] Yllias Chali and **Shafiq Joty**. Improving the Performance of the Random Walk Model for Answering Complex Questions. In *Proceedings of the Association for Computational Linguistics: Human Language Technologies (**ACL-HLT**)* 2008. OH, USA. (Short paper).

[11] Yllias Chali and **Shafiq Joty**. Selecting Sentences for Answering Complex Questions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP**)* 2008, Hawaii, USA.

[12] Gabriel Murray and **Shafiq Joty** and Giuseppe Carenini and Raymond Ng. The University of British Columbia at TAC 2008. In *Proceedings of the Text Analysis Conference (TAC)* 2008**,** Rochester, NY.

[13] Yllias Chali and **Shafiq Joty** and Sheikh Sadid-Al-Hasan. UofL at TAC 2008 Update Summarization and Question Answering. In *Proceedings of the Text Analysis Conference (TAC)* 2008*,* Rochester, NY.

[14] Yllias Chali and **Shafiq Joty**. Exploiting Syntactic and Shallow Semantic Kernels to Improve Random Walks for Complex Question Answering. In *Proceedings of the 20th IEEE Int'l Conference on Tools with Artificial Intelligence (ICTAI)* 2008. Dayton, OH, November, IEEE.

[15] Yllias Chali and **Shafiq Joty**. Answering Complex Questions Using Query-focused Summarization Technique. In *Proceedings of the 20th IEEE Int'l Conference on Tools with Artificial Intelligence (ICTAI)* 2008. Dayton, OH, November, IEEE. (Short paper).

[16] Yllias Chali and **Shafiq Joty**. Unsupervised Approach for Selecting Sentences in Query-based Summarization. In *Proceedings of the 21st FLAIRS conference (FLAIRS)* 2007, Coconut Grove, Florida,

[17] Yllias Chali and **Shafiq Joty**. University of Lethbridge's Participation in DUC-2007 Main Task. In *Proceedings of the Document Understanding Conference (DUC)* 2007*,* Rochester, NY, NIST.

[18] Yllias Chali and **Shafiq Joty**. University of Lethbridge's Participation in TREC-2007 Main Task**.** In *Proceedings of the Sixteenth Text REtrieval Conference (TREC)* 2007, Gaithersburg, 2007, NIST.

[19] Yllias Chali and **Shafiq Joty.** Word Sense Disambiguation Using Lexical Cohesion**.** In *Proceedings of the 4th International Conference on Semantic Evaluations (SemEval)* 2007, Prague, ACL.

Refereed Workshop Publications

[1] **Shafiq Joty,** Giuseppe Carenini and Raymond Ng. Automatic Topic Labeling in Asynchronous Conversations. *The Pacific Northwest Regional NLP Workshop (NW-NLP)* 2012, Microsoft Research, Redmond.

[2] Wei Jin, **Shafiq Joty,** Giuseppe Carenini and Raymond Ng. Detecting Informative Blog Comments using Tree Structured Conditional Random Fields. *The Pacific Northwest Regional NLP Workshop (NW-NLP)* 2012, Microsoft Research, Redmond.

[3] **Shafiq Joty,** Giuseppe Carenini, Gabriel Murray and Raymond Ng. Exploiting Conversation Features for Finding Topics in Emails. *The Pacific Northwest Regional NLP Workshop (NW-NLP)* 2010. Microsoft Research, Redmond.

[4] **Shafiq Joty,** Giuseppe Carenini, Gabriel Murray and Raymond Ng. Finding Topics in Emails: Is LDA enough? *The NIPS workshop on applications for topic models: text and beyond* 2009. Whistler, B.C. Canada.

Technical Skills

- **Operating System experience:** Windows, Linux.
- **Programming Languages:** Python, Perl, C/C++, Matlab, Java, Shell Scripting.

References

Dr. Giuseppe Carenini
Dept. of Computer Science
University of British Columbia
2366 Main Mall, Vancouver, B.C.
Canada V6T 1Z4
Phone: (604) 822-5109
Email: carenini@cs.ubc.ca

Dr. Raymond T. Ng
Dept. of Computer Science
University of British Columbia
2366 Main Mall, Vancouver, B.C.
Canada V6T 1Z4
Phone: (604) 822-2394
Email: rng@cs.ubc.ca