

**LIBR 559N: TEXT ANALYTICS – Course Syllabus (3)<sup>1</sup>**  
**Muhammad Abdul-Mageed**  
**The School of Library, Archival and Information Studies (SLAIS)**  
**The University of British Columbia**

**Course title:** TEXT ANALYTICS

**Program:** MLIS, Dual

**Year:** 2016, Winter Session I

**Time:** Thurs. 8:00 –11:00 a.m.

**Location:** Room 460

**Instructor:** Dr. Muhammad Abdul-Mageed

**Office location:** TBA

**Office phone:** TBA

**Office hours:** By appointment

**E-mail address:** [mabdulma@umail.iu.edu](mailto:mabdulma@umail.iu.edu) (Subject to change)

**iSchool@UBC Student Portal:** <http://connect.ubc.ca>

## **1. Goals:**

The goals of this course are

- To study the nature of text as a data source for knowledge discovery and identify its relevance to the information needs of diverse individuals, communities and organizations. [1.1]
- To study some of the techniques by which text is automatically processed.
- Collaborate effectively with peers through course assignments. [3.1]
- To demonstrate the types of information which can be extracted from text, and the applications of these types.
- To examine the tools which support various types of text processing and analysis and apply them to address information needs, questions, and issues. [4.1]

## **2. Course Objectives:**

Upon completion of this course students will be able to:

- Discuss the various ways in which text can be analyzed, and appropriate uses of each
- Use open source text analytic tools
- Develop simple text analysis tools
- Design a project for textual analysis suitable for a specific domain

---

<sup>1</sup> The syllabus is subject to minor changes.

### 3. Course Topics:

- Overview
- Methods & Approaches
  - ✓ Content Analysis
  - ✓ Natural Language Processing
  - ✓ Clustering & Topic Detection
  - ✓ Simple Predictive Modeling
- Applications
  - ✓ Sentiment Analysis
  - ✓ Emotion Detection
  - ✓ Scholarly Communication
  - ✓ Health
  - ✓ Visualization

### 4. Prerequisites:

- MLIS and Dual MAS/MLIS: LIBR 500, LIBR 501, LIBR 502
- MAS: completion of MAS core and permission of the SLAIS Graduate Adviser
- Access to a computer: There will be machines in the lab where class is held, but you will need to use your own machine or have access to a machine on a regular basis. You should make your own arrangements for this.

### 5. Format of the course:

- This course will involve lectures, class hands-on activities, individual and group work, and instructor-, peer-, and self-assessment.

### 6. Course syllabus:

#### **Books:**

(Only chosen chapters from the books will be assigned. Specific chapters are listed in the “Book Chapters” section below. Soft copies of the book chapters will be made available to students):

- Abdul-Mageed, M. (2016). Sentiment Analysis. [Draft document to be shared with class]
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media, Inc. [[link](#)]
- Bayley, R., Cameron, R., & Lucas, C. (2013). *The Oxford handbook of sociolinguistics*. Oxford University Press.
- Gaskell, M. G. (2007). *The Oxford handbook of psycholinguistics*. Oxford University Press.
- Huang, R. (2017). *The Oxford handbook of pragmatics*. Oxford University Press.
- Jurafsky, D. & Martin, J. H. (2000). *Speech & language processing*. Pearson.
- Manning, C. D., Raghavan, P., & H. Schütze. (2008). Introduction to information retrieval. Cambridge University Press.
- Mitkov, R. (2005). *The Oxford handbook of computational linguistics*. Oxford University Press.
- Taylor, J. R. (Ed.). (2015). *The Oxford handbook of the word*. Oxford University Press.

### Book Chapters:

- Burrige, K. (2015). Taboo Words. In *The Oxford Handbook of the Word*.
- Crystal, D. (2015). The Lure of Words. In *The Oxford Handbook of the Word*.
- Fellbaum, C. (2015). Lexical Relations. In *The Oxford Handbook of the Word*.
- Goddard, C. (2015). Words as carriers of cultural meaning. In *The Oxford handbook of the word*.
- Hoey, M. (2015). Words and Their Neighbours. In *The Oxford Handbook of the Word*.
- Huang, R. (2017). What is Pragmatics? *The Oxford Handbook of Pragmatics*.
- Loudermilk, B. C. (2013). Psycholinguistic approaches. In *The Oxford Handbook of Psycholinguistics*.
- Levinson, S. (2017). Speech Acts. *The Oxford Handbook of Pragmatics*.
- Moon, R. (2015). Multi-word Items. In *The Oxford Handbook of the Word*.
- Ramsay, A. (2005). Discourse. *The Oxford handbook of computational linguistics*. Oxford University Press.
- Raskin, V. (2015). Funny Words: Verbal Humour. In *The Oxford Handbook of the Word*.
- Walker, J. A., & Meyerhoff, M. (2013). Studies of the community and the individual. *The Oxford Handbook of Sociolinguistics*, 175.

### Articles:

- Abdul-Mageed, M. M., & Herring, S. C. (2008). Arabic and English news coverage on aljazeera.net. [\[pdf\]](#)
- Herring, S. C. (2010). Web content analysis: Expanding the paradigm. In J. Hunsinger, M. Allen, & L. Klawns (Eds.), *The International Handbook of Internet Research* (pp. 233-249). Berlin: Springer Verlag. [\[pdf\]](#)
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132.
- Budak, C., Goel, S., & Rao, J. M. (2014). Fair and balanced? quantifying media bias through crowdsourced content analysis. *Quantifying Media Bias Through Crowdsourced Content Analysis (November 17, 2014)*. [\[pdf\]](#)
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39-41. [\[pdf\]](#)
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International journal of lexicography*, 3(4), 235-244. [\[pdf\]](#)
- Blei et al. (2003) (optional)
- Lane (2015) Ch14 sections A-C: [Intro. to linear regression](#).
- Abdul-Mageed, M. & Diab, M. (2014). SANA: A Large Scale, Multi-Genre, Multi-Dialect Lexicon for Arabic Sentiment Analysis. *The 9th International Conference on Language Resources and Evaluation (LREC2014)*, May 26-31, Reykjavik, Iceland. [\[pdf\]](#)
- Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC* (Vol. 6, pp. 417-422). [\[pdf\]](#)
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May. *European Language Resources Association (ELRA)*. [\[pdf\]](#)
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307. [\[pdf\]](#)
- Nakov, P., Rosenthal, S., Kiritchenko, S., Mohammad, S. M., Kozareva, Z., Ritter, A., ... & Zhu, X. (2016). Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts. *Language Resources and Evaluation*, 50(1), 35-65. [\[pdf\]](#)
- Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. *Handbook of affective sciences*, 572, V595. [\[pdf\]](#)

- Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, 5(2), 119-124. [[pdf](#)]
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, 110(1), 145. [[pdf](#)]
- Saima Aman and Stan Szpakowicz. 2007. Identifying expressions of emotion in text. In Vclav Matoušek and Pavel Mautner, editors, *Text, Speech and Dialogue*, volume 4629 of Lecture Notes in Computer Science, pages 196-205. Springer Berlin / Heidelberg. [[pdf](#)]
- Strapparava, C., & Mihalcea, R. (2007, June). Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations* (pp. 70-74). Association for Computational Linguistics. [[pdf](#)]
- Mohammad, S. M. (2012, June). # Emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation* (pp. 246-255). Association for Computational Linguistics. [[pdf](#)]
- Yan, J. L. S., Turtle, H. R., & Liddy, E. D. (2016). EmoTweet-28: A Fine-Grained Emotion Corpus for Sentiment Analysis. *LREC*. [[pdf](#)]
- Yan, J. L. S., & Turtle, H. R. (2016). Exposing a Set of Fine-Grained Emotion Categories from Tweets. In *25th International Joint Conference on Artificial Intelligence* (p. 8). [[pdf](#)]
- Yan, J. L. S., & Turtle, H. R. (2016, June). Exploring Fine-Grained Emotion Detection in Tweets. In *Proceedings of NAACL-HLT* (pp. 73-80). [[pdf](#)]
- De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013, July). Predicting Depression via Social Media. In *ICWSM* (p. 2). [[pdf](#)]
- McIver, D. J., Hawkins, J. B., Chunara, R., Chatterjee, A. K., Bhandari, A., Fitzgerald, T. P., ... & Brownstein, J. S. (2015). Characterizing sleep issues using twitter. *Journal of medical Internet research*, 17(6). [[pdf](#)]
- Preotiuc-Pietro, D., Eichstaedt, J., Park, G., Sap, M., Smith, L., Tobolsky, V., ... & Ungar, L. (2015). The role of personality, age and gender in tweeting about mental illnesses. *NAACL HLT 2015*, 21. [[pdf](#)]
- Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., & Bierut, L. J. (2016). A content analysis of depression-related tweets. *Computers in human behavior*, 54, 351-357.
- Schwartz, H. A., Park, G., Sap, M., Weingarten, E., Eichstaedt, J., Kern, M., ... & Ungar, L. (2015). Extracting Human Temporal Orientation in Facebook Language. In *Proceedings of the The 2015 Conference of the North American Chapter of the Association for Computational Linguistics-Human Language Technologies, NAACL*. [[pdf](#)]
- Park, G., Schwartz, H. A., Sap, M., Kern, M. L., Weingarten, E., Eichstaedt, J. C., ... & Seligman, M. E. (2016). Living in the Past, Present, and Future: Measuring Temporal Orientation With Language. *Journal of personality*. [[pdf](#)]
- Nualart-Vilaplana, J., Pérez-Montoro, M., & Whitelaw, M. (2014). How we draw texts: a review of approaches to text visualization and exploration. *El profesional de la información*, 23(3), 221-235. [[pdf](#)]
- Kucher, K., & Kerren, A. (2015, April). Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific Visualization Symposium (PacificVis)* (pp. 117-121). IEEE.
- Larsen, M. E., Boonstra, T. W., Batterham, P. J., O'Dea, B., Paris, C., & Christensen, H. (2015). We Feel: mapping emotion on Twitter. *IEEE journal of biomedical and health informatics*, 19(4), 1246-1252.
- Kim, K., & Lee, J. (2014). Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. *Pattern Recognition*, 47(2), 758-768.
- Hoque, E., & Carenini, G. (2014, June). ConVis: A visual text analytic system for exploring blog conversations. In *Computer Graphics Forum* (Vol. 33, No. 3, pp. 221-230). [[pdf](#)]
- Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600), 453-458.

- Polonski, V. (2016, August, 14). The biggest threat to democracy? Your social media feed. *World Economic Forum*. Retrieved from: [https://www.weforum.org/agenda/2016/08/the-biggest-threat-to-democracy-your-social-media-feed?utm\\_content=bufferd2b2f&utm\\_medium=social&utm\\_source=facebook.com&utm\\_campaign=buffer](https://www.weforum.org/agenda/2016/08/the-biggest-threat-to-democracy-your-social-media-feed?utm_content=bufferd2b2f&utm_medium=social&utm_source=facebook.com&utm_campaign=buffer).
- Sample, I. (2016, April, 2016). Neuroscientists create ‘atlas’ showing how words are organised in the brain. *The Guardian*. Retrieved from: <https://www.theguardian.com/science/2016/apr/27/brain-atlas-showing-how-words-are-organised-neuroscience>. **Nature video:** <https://www.youtube.com/watch?v=k61nJkx5aDQ>.
- Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Van Harmelen, A. L., ... & Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior research methods*, 45(1), 169-177. [[pdf](#)]
- Yin, D., Bond, S., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *Mis Quarterly*, 38(2), 539-560. [[pdf](#)]
- Wondra, J. D., & Ellsworth, P. C. (2015). An appraisal theory of empathy and other vicarious emotional experiences. *Psychological review*, 122(3), 411. <http://www.apa.org/pubs/journals/features/rev-a0039252.pdf>
- Winter, B. (2016). The Sensory Structure of the English Lexicon. [Ph.D. dissertation: downloaded]
- Mohammad, S. M., Zhu, X., Kiritchenko, S., & Martin, J. (2015). Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management*, 51(4), 480-499. [[pdf](#)]
- Mohammad, S. M., Sobhani, P., & Kiritchenko, S. (2016). Stance and sentiment in tweets. *arXiv preprint arXiv:1605.01655*. [[pdf](#)]
- Winter, B. (2016). Taste and smell words form an affectively loaded and emotionally flexible part of the English lexicon. *Language, Cognition and Neuroscience*, 1-14. [[pdf](#)]
- Citron, F. M., Güsten, J., Michaelis, N., & Goldberg, A. E. (2016). Conventional metaphors in longer passages evoke affective brain response. *NeuroImage*, 139, 218-230. [[pdf](#)]
- Milnea, D., Parisb, C., Christensenc, H., Batterhamc, P., & O’Deac, B. (2015, August). We Feel: Taking the emotional pulse of the world. In *Proceedings 19th Triennial Congress of the IEA* (Vol. 9, p. 14). [[pdf](#)]
- Citron, F. M., Gray, M. A., Critchley, H. D., Weekes, B. S., & Ferstl, E. C. (2014). Emotional valence and arousal affect reading in an interactive way: neuroimaging evidence for an approach-withdrawal framework. *Neuropsychologia*, 56, 79-89.
- Wikipedia articles: [syntax](#), [semantics](#), [pragmatics](#), [discourse analysis](#), [sociolinguistics](#).

### **Tools:**

- [TweetNLP](#)
- [NLTK Tagging](#)
- [Gensim lda topics](#)
- [AntConc](#) concordance, [manual](#)

### **Datasets & Resources:**

Students will be pointed to several available datasets and some data collected by the instructor will be also shared. Sample resources include:

- [WordNet](#)
- [SentiWordNet](#)
- [Wikipedia dumps](#)
- [Literary datasets](#)
- Twitter datasets
- ... ..

## 7. Calendar / Weekly schedule and readings (tentative)

BKL=Bird, Klein, & Loper book

J&M= Jurafsky & Martin

MRS= Manning, Raghavan & Schütze

Week	Date	Topic	Readings
1	Sept. 8	Course Overview  Introduction to Text Analytics	<ul style="list-style-type: none"> <li>Crystal, D. (2015). The Lure of Words. In <i>The Oxford Handbook of the Word</i>.</li> <li>Wikipedia articles: <a href="#">syntax</a>, <a href="#">semantics</a>, <a href="#">pragmatics</a>, <a href="#">discourse analysis</a>, <a href="#">sociolinguistics</a></li> <li>Ramsay, A. (2005). Discourse. <i>The Oxford handbook of computational linguistics</i>. Oxford University Press.</li> </ul>
2	Sept. 15	Nature of Text  Culture  Concordance	<ul style="list-style-type: none"> <li>Moon, R. (2015). Multi-word Items. In <i>The Oxford Handbook of the Word</i>.</li> <li>Hoey, M. (2015). Words and Their Neighbours. In <i>The Oxford Handbook of the Word</i>.</li> <li>Burridge, K. (2015). Taboo Words. In <i>The Oxford Handbook of the Word</i>.</li> <li>Goddard, C. (2015). Words as carriers of cultural meaning. In <i>The Oxford handbook of the word</i>.</li> <li>Raskin, V. (2015). Funny Words: Verbal Humour. In <i>The Oxford Handbook of the Word</i>.</li> <li>Loudermilk, B. C. (2013).</li> <li><a href="#">AntConc</a> concordance, <a href="#">manual</a></li> </ul>
3	Sept. 22	Text & Interaction	<ul style="list-style-type: none"> <li>Psycholinguistic approaches. In <i>The Oxford Handbook of Psycholinguistics</i>.</li> <li>Walker, J. A., &amp; Meyerhoff, M. (2013). Studies of the community and the individual. <i>The Oxford Handbook of Sociolinguistics</i>, 175.</li> <li>Huang, R. (2016). What is Pragmatics? <i>The Oxford Handbook of Pragmatics</i>.</li> <li>Levinson, S. (2016). Speech Acts. <i>The Oxford Handbook of Pragmatics</i>.</li> </ul>
<b>Methods &amp; Approaches:</b>			
4	Sept. 29	Content Analysis (of News Discourse)	<ul style="list-style-type: none"> <li>Content analysis: <a href="#">Herring (2010)</a></li> <li><a href="#">Abdul-Mageed &amp; Herring (2008)</a></li> <li>Bakshy, E., Messing, S., &amp; Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. <i>Science</i>, 348(6239), 1130-1132.</li> <li>Budak, C., Goel, S., &amp; Rao, J. M. (2014). Fair and balanced? quantifying media bias through crowdsourced content analysis. <i>Quantifying Media Bias Through Crowdsourced Content Analysis (November 17, 2014)</i>. <a href="#">[pdf]</a></li> </ul>

5	Oct. 6	Natural Language Processing	<ul style="list-style-type: none"> <li>J&amp;M: <a href="#">Ch01</a>, <a href="#">Ch02</a></li> </ul>
6	Oct. 13	Frequency Analysis Co-Occurrence & N-grams POS	<ul style="list-style-type: none"> <li>MRS: <a href="#">Ch02</a> (2.0-2.2)</li> <li>J&amp;M: <a href="#">Ch09</a></li> <li><a href="#">TweetNLP</a></li> <li><a href="#">NLTK Tagging</a></li> </ul>
7	Oct. 20	Lexical Relations & Topics	<ul style="list-style-type: none"> <li><a href="#">WordNet</a></li> <li>Fellbaum, C. (2015). Lexical Relations. In <i>The Oxford Handbook of the Word</i>.</li> <li>Miller (1995); Miller et al. (1990) (<b>optional</b>)</li> <li>Blei et al. (2003) (<b>optional</b>)</li> <li><a href="#">Gensim lda topics</a></li> </ul>
8	Oct. 27	Clustering: K-Nearest Neighbors Simple Predictive Modeling: Regression	<ul style="list-style-type: none"> <li>MRS: <a href="#">Ch06</a></li> <li>Lane (2015) Ch14 sections A-C: <a href="#">Intro. to linear regression</a>: (<b>Optional</b>)</li> </ul>
		<b>Applications:</b> <b><i>Sentiment</i></b>	
9	Nov. 3	Introduction to Sentiment Analysis	<ul style="list-style-type: none"> <li>MAM: Ch01</li> <li>BL: Ch01</li> </ul>
10	Nov. 10	Learning Prior Polarity Sentiment Prediction	<ul style="list-style-type: none"> <li>MAM: Ch02</li> <li>BL Ch02 (<b>Optional</b>)</li> <li><a href="#">SentiWordNet</a></li> <li>Abdul-Mageed, M. &amp; Diab, M. (2014). SANA: A Large Scale, Multi-Genre, Multi-Dialect Lexicon for Arabic Sentiment Analysis. <i>The 9th International Conference on Language Resources and Evaluation (LREC2014)</i>, May 26-31, Reykjavik, Iceland. <a href="#">[pdf]</a></li> <li>Esuli, A., &amp; Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In <i>Proceedings of LREC</i> (Vol. 6, pp. 417-422). <a href="#">[pdf]</a> (<b>Optional</b>)</li> <li>Baccianella, S., Esuli, A., &amp; Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In <i>Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)</i>, Valletta, Malta, May. <i>European Language Resources Association (ELRA)</i>. <a href="#">[pdf]</a></li> <li>Taboada, M., Brooke, J., Tofiloski, M., Voll, K., &amp; Stede, M. (2011). Lexicon-based methods for sentiment analysis. <i>Computational linguistics</i>, 37(2), 267-307. <a href="#">[pdf]</a></li> <li>Nakov, P., Rosenthal, S., Kiritchenko, S., Mohammad, S. M., Kozareva, Z., Ritter, A., ...</li> </ul>

			<p>&amp; Zhu, X. (2016). Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts. <i>Language Resources and Evaluation</i>, 50(1), 35-65. <a href="#">[pdf]</a></p> <p><b>(Optional)</b></p>
<b>Emotion</b>			
11	Nov. 17	Theory	<ul style="list-style-type: none"> <li>• Ellsworth, P. C., &amp; Scherer, K. R. (2003). Appraisal processes in emotion. <i>Handbook of affective sciences</i>, 572, V595. <a href="#">[pdf]</a></li> <li>• Moors, A., Ellsworth, P. C., Scherer, K. R., &amp; Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. <i>Emotion Review</i>, 5(2), 119-124. <a href="#">[pdf]</a></li> <li>• Russell, J. A. (2003). Core affect and the psychological construction of emotion. <i>Psychological review</i>, 110(1), 145. <a href="#">[pdf]</a></li> </ul>
12	Nov. 24	Emotion Annotation & Prediction	<ul style="list-style-type: none"> <li>• Saima Aman and Stan Szpakowicz. 2007. Identifying expressions of emotion in text. In Vclav Matoušek and Pavel Mautner, editors, <i>Text, Speech and Dialogue</i>, volume 4629 of Lecture Notes in Computer Science, pages 196-205. Springer Berlin / Heidelberg. <a href="#">[pdf]</a></li> <li>• Strapparava, C., &amp; Mihalcea, R. (2007, June). Semeval-2007 task 14: Affective text. In <i>Proceedings of the 4th International Workshop on Semantic Evaluations</i> (pp. 70-74). Association for Computational Linguistics. <a href="#">[pdf]</a></li> <li>• <b>(Optional)</b></li> <li>• Mohammad, S. M. (2012, June). # Emotional tweets. In <i>Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation</i> (pp. 246-255). Association for Computational Linguistics. <a href="#">[pdf]</a></li> <li>• Yan, J. L. S., Turtle, H. R., &amp; Liddy, E. D. (2016). EmoTweet-28: A Fine-Grained Emotion Corpus for Sentiment Analysis. <i>LREC</i>. <a href="#">[pdf]</a></li> <li>• Yan, J. L. S., &amp; Turtle, H. R. (2016). Exposing a Set of Fine-Grained Emotion Categories from Tweets. In <i>25th International Joint Conference on Artificial Intelligence</i> (p. 8). <a href="#">[pdf]</a> <b>(Optional)</b></li> <li>• Yan, J. L. S., &amp; Turtle, H. R. (2016, June). Exploring Fine-Grained Emotion Detection in Tweets. In <i>Proceedings of NAACL-HLT</i> (pp. 73-80). <a href="#">[pdf]</a></li> </ul>
13	Dec. 1	Scholarly Communication, Health, Visualization	<ul style="list-style-type: none"> <li>• Abdul-Mageed (In Preparation). Knowledge Negotiation in Wikipedia.</li> <li>• De Choudhury, M., Gamon, M., Counts, S., &amp; Horvitz, E. (2013, July). Predicting Depression</li> </ul>



			<p>via Social Media. In <i>ICWSM</i> (p. 2). <a href="#">[pdf]</a></p> <ul style="list-style-type: none"> <li>• McIver, D. J., Hawkins, J. B., Chunara, R., Chatterjee, A. K., Bhandari, A., Fitzgerald, T. P., ... &amp; Brownstein, J. S. (2015). Characterizing sleep issues using twitter. <i>Journal of medical Internet research</i>, 17(6). <a href="#">[pdf]</a></li> <li>• Preotiuc-Pietro, D., Eichstaedt, J., Park, G., Sap, M., Smith, L., Tobolsky, V., ... &amp; Ungar, L. (2015). The role of personality, age and gender in tweeting about mental illnesses. <i>NAACL HLT 2015</i>, 21. <a href="#">[pdf]</a></li> <li>• Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., &amp; Bierut, L. J. (2016). A content analysis of depression-related tweets. <i>Computers in human behavior</i>, 54, 351-357.</li> <li>• Nualart-Vilaplana, J., Pérez-Montoro, M., &amp; Whitelaw, M. (2014). How we draw texts: a review of approaches to text visualization and exploration. <i>El profesional de la información</i>, 23(3), 221-235. <a href="#">[pdf]</a></li> <li>• Kucher, K., &amp; Kerren, A. (2015, April). Text visualization techniques: Taxonomy, visual survey, and community insights. In <i>2015 IEEE Pacific Visualization Symposium (PacificVis)</i> (pp. 117-121). IEEE.</li> <li>• Hoque, E., &amp; Carenini, G. (2014, June). ConVis: A visual text analytic system for exploring blog conversations. In <i>Computer Graphics Forum</i> (Vol. 33, No. 3, pp. 221-230). <a href="#">[pdf]</a></li> </ul>
14	Dec. 8	Poster presentations	

## 8. Course Assignments:

<i>Assignment</i>	<i>Due date</i>	<i>Weight</i>
ASSIGNMENT 1: Class presentation		10%
ASSIGNMENT 2: Personal readings journal	Throughout first 10 weeks	10%
ASSIGNMENT 3: First analytics task— <i>language analysis</i>	Oct. 20	15%
ASSIGNMENT 4: Second analytics task— <i>text mining/automated content analysis</i>	Nov. 3	20%
ASSIGNMENT 5: Term Project (GROUP)		
Proposal	Nov. 17	5%
Poster	Dec. 8	10%
Report/Product	Dec. 10	30%

**\* Notes about assignments:**

- All written assignments must be submitted before 11:55 pm of a respective assignment date;
- Project proposal/outline must be in a pdf format;
- All submitted files should be labeled with your last name(s) followed by an underscore and an assignment code. Assignment codes are two digit numbers, e.g., “assignment01,” “assignment02,” etc.
  - For example: “abdul-mageed\_assignment01.pdf”

**Assignments in Detail**

***Assignment 1: Class presentation***

Each student is expected to present once during the course of the semester. For this assignment, each student will choose one or more papers and/or book chapters in consultation with the instructor. The student will then be responsible for preparing slides (e.g., Microsoft PowerPoint or pdf format) for the content and presenting to class in one of the class sessions. The following criteria are required for the presentation:

- Students are expected to submit their slides to the instructor by 5:00 pm the day before the presentation is due. The instructor will then make the slides available to the whole class.
- Students are encouraged to rehearse two-to-three times before presenting in class, make use of body language, use illustrative examples and visualizations, and deliver their presentation in the assigned time.

**Grading:** Grading for this assignment will be based on content, presentation skills, clarity of presentation, and ability to answer questions.

**Deliverables: Presentation slides and actual presentation**

Students are expected to submit their slides to instructor and deliver the presentation in class to satisfy this requirement.

***Assignment 2: Personal readings journal***

Students are required to keep a personal readings journal where they summarize and reflect on the weekly readings. The journal should be a single Word document that students update with the new content every week. Each week’s log can be in the range of 300-400 words summarizing the main ideas of the readings. Many of the papers we read describe the nature of text as a source of information and discuss different analytics. Students are expected to summarize the main points the author(s) of each paper/chapter make(s). For example, for a sentiment analysis task, students are expected to identify how the concept of sentiment is operationalized, the basic idea of a method used, the datasets exploited, and the results acquired. In addition, students are expected to engage critically by possibly identifying limitations and extensions of the work they can think of. The personal readings journal is meant to keep students involved throughout the course of the term. Students are also expected to participate in discussions in class, based on their readings and the journal should help with this regard. Students are expected to have a total of 10 logs in the journal, corresponding to the readings of the first 10 weeks. Each week’s log is only accepted if the student attends class for a respective week. If a student misses a class, this automatically means he/she is not allowed to update/submit a log for the session/sessions he/she missed. As

such, failure to attend class without a necessary excuse (e.g., a sickness or another emergency) will likely result in a lower class grade.

Deliverables: A personal readings journal

A total of 10 logs in a single Word document.

**Assignment 3: First analytics task—Language analysis**

For this assignment, students are expected to analyze the *language* of a category of text (e.g., literary, academic, social media) of their selection using a rubric provided by the instructor. Several datasets will be suggested by the instructor and students can choose one or more of these datasets or choose and/or construct one or more of their own. The rubric will guide students for analyzing the language of the dataset(s) in terms of, for example, word frequency and length, parts of speech (e.g., nouns, verbs, adjectives), word category (e.g., softeners/hedges and self-mentions in academic writing, words expressing happiness), topics as identified by word groups (e.g., family topic, recreation topic, crime topic), phrases and multi-word expressions (e.g., collocations, idioms, metaphors). Based on the analysis, students are expected to provide a critical description of the texts involved and possibly distinguishing them from other texts and/or uncovering relationships or concepts communicated by the text authors (e.g., how academicians negotiate knowledge creation and promote self, differences between characters in literary works). As part of the language analysis, students should include a section describing how they this type of analysis can help in our overall understanding of a given concept and/or field.

Deliverables: A written report (~ 2000 words)

For this assignment, students are required to provide a written report with their *language* analysis. The report should be supported by quantitative evidence as well as a qualitative appraisal of any uncovered or meaningful relationships within the text. Students should include a brief introduction, follow by the analysis itself (students have the freedom to thematically organize the analysis into different subsections), a section describing the potential of language analysis and its utility as described above, and a conclusion. Students are expected to follow norms of academic writing and document and support their claims by academic and professional references as appropriate. Students are also encouraged to make use of tables, figures, screenshots, visualizations, etc.

**Assignment 4: Second analytics task—text mining/automated content analysis**

For this assignment, students are required to work with a dataset provided by the instructor on a mining task. The mining task will be sentiment analysis or emotion detection. It is also possible for students to agree with the instructor on another task provided that students clearly define the task and guarantee access to necessary data. In either case, students will be required to use off-the-shelf software and/or code of their own to detect sentiment/emotion in the data and write a description of the methods they use and the results. The instructor will provide a detailed description of the steps students will need to follow to articulate the task, including possible use of lexical resources and software and methods.

The following criteria is required for this assignment:

- If students choose to write their own code, they must run the code before submitting, and ensure it is working. Also, should should provide the output of the code itself. For code that writes to a file, students are required to deliver the output files as well.
- For any off-the-shelf software used, a description of the software and its functionality should be provided.

- For each sub-part of the task, clear description should be provided as to how it is articulated. If code is written, students provide short and to-the-point comments in the code.

Deliverables: Written description (~ 2000 words) + code if any

- The written description should describe the task, the methods used, provide a brief literature review (~ 5-7 sources), results, and a discussion and conclusion section.
- Any accompanying code should be provided as part of the deliverable.
- Students are required to follow the norms of academic writing as with Assignment 3.

***Assignment 5: Term Project***

The purposes of this assignment include:

- Identifying, analyzing, assessing, and solving a problem via use of text data;
- Applying text analytics methods to a practical task of the student's choice, after consultation with the instructor;
- Developing oral and written communication skills through discussions with classmates and instructor;
- Demonstrating ability to work as part of a team, including initiative taking, integrity, dependability and co-operation, and effective collaboration.

For this assignment, each student is required to work as part of a group of 2/3\* on a project involving a practical text analytics task. Example projects include sentiment analysis, emotion detection, mining literary work, analyzing academic databases of abstracts or journal papers, large-scale visualization of texts in a meaningful context, etc.

\* A group of a different size may be possible after consultation with the instructor.

**Deliverables**

Proposal (500-700 words)

- Who are the the group members?
- What are you analyzing/mining?
- What motivates your work? Why is it important or useful to undertake the chosen task?
- How does your project compare to other projects people have conducted in the past?
- What are the different steps you will take to ensure success of the project? What are the smaller segments of which the bigger analytics task is composed? And how will you conduct each small task?
- How does the work break down and what each member of the team be contributing?
- Timeline for completing the project, including goals for each segment.

Poster

Each team will prepare a poster to present their final project to the class. Each student is required to participate in the poster preparation and the poster session. Your poster should first motivate your work, explaining its significance and audience, and how it compares to previous research. The team should then explain the various steps the work involved and the final outcome and its functionality. More details about scheduling and poster session place will be provided and discussed toward the end of the semester.

Final Report/Paper (4000-6000 words)

The final deliverable should include:

- A detailed and clear description of your project, including the necessary sections, as appropriate. For example, you will need to include an abstract, introduction, research questions, a literature review, a description of datasets, implementation details (or description of software used) and methods employed, results, and a conclusion involving limitations and future directions;
  - All relevant code, if any;
  - All data used, whenever possible;
  - Pointers to a live version of the project, if any;
  - As appropriate, you should situate your work within the wider context of previous works and approaches, with supporting arguments (~ 15 sources);
  - Employment of figures, tables, and visualizations as appropriate to enhance argument and facilitate communicating your findings/results;
- 

## 9. Course Policies

**Attendance:** The UBC calendar states: “Regular attendance is expected of students in all their classes (including lectures, laboratories, tutorials, seminars, etc.). Students who neglect their academic work and assignments may be excluded from the final examinations. Students who are unavoidably absent because of illness or disability should report to their instructors on return to classes.”

**Evaluation:** All assignments will be marked using the evaluative criteria given in this syllabus and also those provided on the [SLAIS web site](#).

**Written & Spoken English Requirement:** Written and spoken work may receive a lower mark if it is, in the opinion of the instructor, deficient in English.

**Access & Diversity:** Access & Diversity works with the University to create an inclusive living and learning environment in which all students can thrive. The University accommodates students with disabilities who have registered with the Access and Diversity unit: [\[http://www.students.ubc.ca/access/drc.cfm\]](http://www.students.ubc.ca/access/drc.cfm). You must register with the Disability Resource Centre to be granted special accommodations for any on-going conditions.

**Religious Accommodation:** The University accommodates students whose religious obligations conflict with attendance, submitting assignments, or completing scheduled tests and examinations. Please let your instructor know in advance, preferably in the first week of class, if you will require any accommodation on these grounds. Students who plan to be absent for varsity athletics, family obligations, or other similar commitments, cannot assume they will be accommodated, and should discuss their commitments with the instructor before the course drop date. UBC policy on Religious Holidays: <http://www.universitycounsel.ubc.ca/policies/policy65.pdf>.

### Academic Integrity

#### Plagiarism

The Faculty of Arts considers plagiarism to be the most serious academic offence that a student can commit. Regardless of whether or not it was committed intentionally, plagiarism has serious academic consequences and can result in expulsion from the university. Plagiarism involves the improper use of somebody else's words or ideas in one's work.

It is your responsibility to make sure you fully understand what plagiarism is. Many students who think they understand plagiarism do in fact commit what UBC calls "reckless plagiarism." Below is an excerpt on reckless plagiarism from UBC Faculty of Arts' leaflet, "Plagiarism Avoided: Taking Responsibility for Your Work," (<http://www.arts.ubc.ca/arts-students/plagiarism-avoided.html>).

"The bulk of plagiarism falls into this category. Reckless plagiarism is often the result of careless research, poor time management, and a lack of confidence in your own ability to think critically. Examples of reckless plagiarism include:

- Taking phrases, sentences, paragraphs, or statistical findings from a variety of sources and piecing them together into an essay (piecemeal plagiarism);
- Taking the words of another author and failing to note clearly that they are not your own. In other words, you have not put a direct quotation within quotation marks;
- Using statistical findings without acknowledging your source;
- Taking another author's idea, without your own critical analysis, and failing to acknowledge that this idea is not yours;
- Paraphrasing (i.e. rewording or rearranging words so that your work resembles, but does not copy, the original) without acknowledging your source;
- Using footnotes or material quoted in other sources as if they were the results of your own research; and
- Submitting a piece of work with inaccurate text references, sloppy footnotes, or incomplete source (bibliographic) information."

Bear in mind that this is only one example of the different forms of plagiarism. Before preparing for their written assignments, students are strongly encouraged to familiarize themselves with the following source on plagiarism: the Academic Integrity Resource Centre <http://help.library.ubc.ca/researching/academic-integrity>. Additional information is available on the Connect site <http://connect.ubc.ca>.

If after reading these materials you still are unsure about how to properly use sources in your work, please ask me for clarification. Students are held responsible for knowing and following all University regulations regarding academic dishonesty. If a student does not know how to properly cite a source or what constitutes proper use of a source it is the student's personal responsibility to obtain the needed information and to apply it within University guidelines and policies. If evidence of academic dishonesty is found in a course assignment, previously submitted work in this course may be reviewed for possible academic dishonesty and grades modified as appropriate. UBC policy requires that all suspected cases of academic dishonesty must be forwarded to the Dean for possible action.