

# USC Viterbi School of Engineering

## **INF 558/CSCI 563: Building Knowledge Graphs**

**Units: 4**

**Term—Day—Time:**

Fall 2018 – Friday – 2:00 - 5:20pm

**Location:** VKC 150

**Instructor:** Pedro Szekely

**Office:** ISI

**Office Hours:** After each class in VKC 150, or at ISI by appointment

**Contact Info:** szekely@usc.edu, 310-448-8641.

**Instructor:** Jay Pujara

**Office:** RTH 512

**Office Hours:** After each class in VKC 150, or by appointment

**Contact Info:** jpujara@usc.edu, 310-448-8482.

**Teaching Assistant:** Binh Vu

**Office:** TBD

**Office Hours:** Tu, 4:00-5:30pm

**Contact Info:** binhvu@isi.edu

### **Catalogue Course Description**

Foundations, techniques, and algorithms for building knowledge graphs and doing so at scale. Topics include information extraction, data alignment, entity linking, and the Semantic Web.

### **Expanded Course Description**

This course focuses on foundations, techniques, and algorithms for building knowledge graphs. Students will learn the theory and applications of the techniques needed to build and query massive knowledge graphs. Topics include crawling web sites, wrapper learning, information extraction, source alignment, string matching, entity linking, graph databases, querying knowledge graphs, data cleaning, Semantic Web, linked data, graph analytics, and intellectual property. The class will be run as a lecture course with lots of student participation and significant hands-on experience. As an integral part of the course each student will do a project using the research and tools covered in the class.

### **Learning Objectives**

The learning objectives for this course are:

- Understand the algorithms and techniques for crawling web sites, structured data extraction, and information extraction from unstructured text.
- Understand the theory and techniques for cleaning, aligning, matching, and linking data.
- Understand the foundations and techniques of the Semantic Web, including RDF, ontologies, SPARQL, and linked data.
- Understand how to work with graph databases, including how to load massive datasets into such databases, how to organize the data for efficient access, and how to efficiently query the contents.
- Understand the entire process of how to design, construct, and query a knowledge graph to solve real-world problems.

- Understand how to apply the big data tools and infrastructure (e.g., Spark) to build and query knowledge graphs.

### **Required Preparation:**

Prerequisite(s): INF 551 or CSCI 585  
INF 552 or CSCI 567

Recommended Background: Experience programming in Python

### **Course Notes**

The course will be run as a lecture class with student participation strongly encouraged. The first 4-5 weeks of the course are structured as a quickstart to provide a shallow primer on the end-to-end process of knowledge graph construction, followed by deeper presentations and more technical material for the remainder of the course. There are weekly readings and students are encouraged to do the readings prior to the discussion in class. All of the course materials, including the readings, lecture slides, and homeworks will be posted online (<https://bit.ly/2Llz3Cx>). The class project is a significant aspect of this course and at the end of the semester students will present their projects in class.

### **Required Readings and Supplementary Materials**

Required Textbook: none

We use a set of technical papers and book chapters that are all available online. All of the required readings are listed in the course schedule.

### **Description and Assessment of Assignments**

#### **Homework Assignments**

There will be weekly homework assignments for the first 11 weeks of class. The assignments must be done individually. The homework assignments are expected to take 8-10 hours per week. Each assignment is graded on a scale of 0-100 and the specific rubric for each assignment is given in the assignment. The homework topics are listed in the Course Schedule.

#### **Course Project**

An integral part of this course is the course project, which builds on the topics and techniques covered in the class. Students can work in teams of up to two people on this project. They will present their project proposals in class, conduct the project, and then create a video demonstration of the work and present the project in class.

#### *Project Timeline:*

- Week 4: Project proposals presented in class (team members, topic)
- Week 8: Project status update due (1 page status report)
- Week 12: Project status update due (1 page status report)
- Week 15: Project presentation in class (short talk and video demonstration)

*Project description:* Each project team will build a knowledge graph for a topic of their choice. The knowledge graph must combine data from at least 3 different sources and at least 2 of those sites must be from online web sites. The best projects build on many of the topics covered in the class. The homeworks have been designed so that you can work on your projects in the process of doing your homework.

An example project would be to build a knowledge graph of used bicycles that could be purchased near the USC campus. This project would combine data from used sources, such as Craig's List, new bike sources such as BikeNashbar, and bicycle review sites, such as bicycling.com. The project would collect the data from each of these sources using wrapper techniques, extract the details of the used bicycle ads from Craig's List using information extraction techniques, align the data across these various sources to a domain ontology, link the entities across sources to combine the used data with the reviews from bicycling.com and

prices from BikeNashbar, store all of the data into a graph database such as elasticsearch, and then build a simple user interface to show the results by executing queries against the graph database.

*Grading breakdown of the course project:*

- Proposal: 10%
- Project video: 30%
- Presentation: 30%
- Overall project: 30%

## Grading Breakdown

**Quizzes:** There will be weekly quizzes based on the material from the week before. The lowest quiz grade will be dropped. Missed quizzes will receive a zero grade, and there will be no make-up quizzes for any reason.

**Midterm:** There is no mid-term for this class.

**Homework:** There will be weekly homework based on the topics of the class each week.

**Final Exam:** There is a final exam at the end of the semester covering all of the material covered in the class. The final exam will be on the date designated by USC in <https://classes.usc.edu/term-20183/finals/>

**Class Project:** Each student will do a group class project based on the topics covered in the class. Students will propose their own project, do the research and build a proof-of-concept, create a video demonstration of the proof-of-concept, and present the project in class.

Grading Schema:

Quizzes	20%
Homework	20%
Final:	25%
Class Project	35%
<hr/>	
Total	100%

Grades will range from A through F. The following is the breakdown for grading:

94 - 100 = A	74 - 76.9 = C
90 - 93.9 = A-	70 - 73.9 = C-
87 - 89.9 = B+	67 - 69.9 = D+
84 - 86.9 = B	64 - 66.9 = D
80 - 83.9 = B-	60 - 63.9 = D-
77 - 79.9 = C+	Below 60 is an F

## Assignment Submission Policy

Homework assignments are due at 11:59pm on the due date and should be submitted in Blackboard. You can submit homework up to one week late, but you will lose 20% of the possible points for the assignment. After one week, the assignment cannot be submitted.

## Course Schedule: A Weekly Breakdown

	Topics/Daily Activities	Readings	Quizzes & Homeworks	Instructor
<b>Week 1 Aug 24</b>	<b>Course Introduction &amp; Use Case</b>	Pedro Szekely, et al. Building and using a knowledge graph to combat human trafficking. In Proceedings of the 14th	Homework 1: Crawling	Pujara Szekely

	<p><b>Quickstart: Crawling</b></p>	<p>International Semantic Web Conference (ISWC 2015), 2015.  <a href="http://iswc2015.semanticweb.org/sites/iswc2015.semanticweb.org/files/93670175.pdf">http://iswc2015.semanticweb.org/sites/iswc2015.semanticweb.org/files/93670175.pdf</a></p> <p>The Anatomy of a Large Scale Hypertextual Web Search Engine Sergey Brin and Lawrence Page, Seventh International World Wide Web Conference, 1998.  <a href="http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf">http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf</a></p> <p>The Structure of the Web:  <a href="http://science.sciencemag.org/content/sci/294/5548/1849.full.pdf">http://science.sciencemag.org/content/sci/294/5548/1849.full.pdf</a></p> <p>Optional: Web crawling and indexes:  <a href="https://nlp.stanford.edu/IR-book/pdf/20crawl.pdf">https://nlp.stanford.edu/IR-book/pdf/20crawl.pdf</a></p> <p>Optional: Searching the Web_Arvind Arasu et al., ACM Transactions on Internet Technology, 2001  <a href="http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.25.5183">http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.25.5183</a></p>		
<p><b>Week 2 Aug 31</b></p>	<p><b>Quickstart: Information Extraction</b></p>	<p>D. C. Wimalasuriya and D. Dou. <a href="#">Ontology-based Information Extraction: An Introduction and a Survey of Current Approaches</a>. J. Information Science, 36(3), 2010.</p> <p>O. Etzioni, M. Banko, S. Soderland, and D. S. Weld. <a href="#">Open Information Extraction from the Web</a>. Communications of the ACM, 51(12):68–74, 2008.</p> <p><b>Optional:</b></p> <p>J. Zhu, Z. Nie, X. Liu, B. Zhang, and J.-R. Wen. <a href="#">Statsnowball: A Statistical Approach to Extracting Entity Relationships</a>. In Proceedings of the 18th International Conference on World Wide Web, pages 101–110. ACM, 2009.</p> <p>S. Krause, H. Li, H. Uszkoreit, and F. Xu. <a href="#">Large-Scale Learning of Relation-Extraction Rules with Distant Supervision from the Web</a>.</p>	<p>Quiz 1</p> <p>Homework 2: Extracting information from web documents</p>	<p>Szekely</p>

		<p>International Semantic Web Conference, pages 263–278, 2012.</p> <p>Ion Muslea, Steve Minton, and Craig A. Knoblock. A hierarchical approach to wrapper induction. In Proceedings of the 3rd International Conference on Autonomous Agents, Seattle, WA, 1999.  <a href="http://www.isi.edu/integration/papers/muslea99-agents.pdf">http://www.isi.edu/integration/papers/muslea99-agents.pdf</a></p> <p>AnHai Doan, Alon Y. Halevy, and Zachary G. Ives. Principles of Data Integration, chapter 9. Morgan Kaufmann, 2012.  <a href="http://www.sciencedirect.com/science/book/9780124160446">http://www.sciencedirect.com/science/book/9780124160446</a></p> <p>W. Crescenzi, G. Mecca, and P. Merialdo. RoadRunner. Towards automatic data extraction from large web sites. 2001.  <a href="http://www.vldb.org/conf/2001/P109.pdf">http://www.vldb.org/conf/2001/P109.pdf</a></p> <p>B. Cenk Gazen and Steven Minton. Overview of autofeed: An unsupervised learning system for generating webfeeds. In Proceedings of AAAI, 2006.  <a href="http://www.isi.edu/integration/courses/csci548/Papers/gazen06-aaai.pdf">http://www.isi.edu/integration/courses/csci548/Papers/gazen06-aaai.pdf</a>.</p>		
<p><b>Week 3</b> <b>Sep 7</b></p>	<p><b>Quickstart:</b> <b>Knowledge Representation &amp; Entity Linking</b></p>	<p>A. Barr and J. Davidson. <a href="#">Representation of Knowledge</a>, in Handbook of AI, volume 1, Chapter 3A-B, pages 141–160.</p> <p>W. Cohen, P. Ravikumar, and S. Fienberg. <a href="#">A Comparison of String Distance Metrics for Name-matching Tasks</a>. Conference on Information Integration on the Web, 2003.</p> <p>J. Pujara and L. Getoor. <a href="#">Generic Statistical Relational Entity Resolution in Knowledge Graphs</a>. StaRAI 2016.</p> <p>Optional:</p> <p>I. P. Fellegi and A. B. Sunter. A theory for record linkage. Journal of the American Statistical Association, 64(328):1183–1210, 1969.</p> <p>W. E. Winkler. <a href="#">The state of record linkage and current research problems</a>. In Statistical</p>	<p>Quiz 2</p> <p>Homework 3: Representing knowledge and linking entities</p>	<p>Pujara</p>

		<p>Research Division, US Census Bureau. Citeseer, 1999.</p> <p>G. Navarro. A Guided Tour to Approximate String Matching. ACM Comput. Surv., 33 (1):31–88, Mar. 2001. ISSN 0360-0300. doi: 10.1145/375360.375365. URL <a href="http://doi.acm.org/10.1145/375360.375365">http://doi.acm.org/10.1145/375360.375365</a></p> <p>Matthew Michelson and Craig A. Knoblock. Semantic Annotation of Unstructured and Ungrammatical Text. In Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-2005), Edinburgh, Scotland, 2005. <a href="http://www.isi.edu/integration/papers/michelson05-ijcai.pdf">http://www.isi.edu/integration/papers/michelson05-ijcai.pdf</a></p> <p>Andrew McCallum. Information Extraction: Distilling Structured Data from Unstructured Text . ACM Queue, volume 3, Number 9, November 2005. <a href="http://people.cs.umass.edu/~mccallum/papers/acm-queue-ie.pdf">http://people.cs.umass.edu/~mccallum/papers/acm-queue-ie.pdf</a></p> <p>Bo Wu, Pedro Szekely, and Craig A. Knoblock. Minimizing user effort in transforming data by example. In Proceedings of the International Conference on Intelligent User Interface, 2014. <a href="http://www.isi.edu/integration/papers/wu14-iui.pdf">http://www.isi.edu/integration/papers/wu14-iui.pdf</a>.</p> <p>Open Refine, Explore data. <a href="http://youtu.be/B70J_H_zAWM">http://youtu.be/B70J_H_zAWM</a>.</p> <p>Open Refine, Clean and transform data. <a href="http://youtu.be/cO8NVCs_Ba0">http://youtu.be/cO8NVCs_Ba0</a>.</p> <p>Open Refine, Reconcile and match data. <a href="http://youtu.be/5tsyz3ibYzk">http://youtu.be/5tsyz3ibYzk</a>.</p> <p>G. Antoniou and F. Van Harmelen. A Semantic Web Primer. MIT press, 2004</p> <p>F. van Harmelen, V. Lifschitz, and B. Porter. Handbook of Knowledge Representation. Elsevier Science, San Diego, USA, 2007. ISBN 0444522115, 9780444522115</p>		
--	--	---	--	--

		P. Hitzler, M. Krotzsch, and S. Rudolph. Knowledge Representation for the Semantic Web. KI, 2009		
<b>Week 4 Sep 14</b>	<b>Quickstart: DIG, RDF, &amp; Triplestores</b>	<p>Wrangler: Interactive visual specification of data transformation scripts. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011. <a href="http://vis.stanford.edu/papers/wrangler">http://vis.stanford.edu/papers/wrangler</a>.</p> <p>Frank Manola and Eric Miller. Rdf primer. Technical report, W3C, February 2004. <a href="http://www.w3.org/TR/2004/REC-rdf-primer-20040210/">http://www.w3.org/TR/2004/REC-rdf-primer-20040210/</a></p> <p>Tim Berners-Lee. Why rdf model is different from the xml model. Technical report, W3C, 1998. <a href="http://www.w3.org/DesignIssues/RDF-XML.html">http://www.w3.org/DesignIssues/RDF-XML.html</a>.</p> <p>Rdf vocabulary description language 1.0: Rdf schema. Technical report, W3C, February 2004. <a href="http://www.w3.org/TR/2004/REC-rdf-schema-20040210/">http://www.w3.org/TR/2004/REC-rdf-schema-20040210/</a></p> <p>Ben Adida, Ivan Herman, Manu Sporny, and Mark Birbeck. Rdfa 1.1 primer rich structured data markup for web documents. Technical report, W3C, June 2012. <a href="http://www.w3.org/TR/2012/NOTE-rdfa-primer-20120607/">http://www.w3.org/TR/2012/NOTE-rdfa-primer-20120607/</a></p>	<p>Quiz 3</p> <p>Homework 4: Using DIG</p>	Szekely
<b>Week 5 Sep 21</b>	<b>Quickstart: Querying &amp; Analysis</b>	<p>Oren, E., Delbru, R., Catasta, M., Cyganiak, R., Stenzhorn, H., &amp; Tummarello, G. (2008). Sindice. com: a document-oriented lookup index for open linked data. <i>International Journal of Metadata, Semantics and Ontologies</i>, 3(1), 37-52. <a href="http://wtlab.um.ac.ir/images/e-library/linked_data/other/Sindice.pdf">http://wtlab.um.ac.ir/images/e-library/linked_data/other/Sindice.pdf</a></p> <p>Freitas, A., Oliveira, J. G., Curry, E., O’Riain, S., &amp; da Silva, J. C. P. (2011, June). Treo: combining entity-search, spreading activation and semantic relatedness for querying linked data. In <i>Proc. of 1st Workshop on Question Answering over Linked Data (QALD-1) at the 8th Extended Semantic Web Conference (ESWC 2011)</i>.</p>	<p>Quiz 4</p> <p>Homework 5: KG Analysis</p>	Pujara

		<p><a href="https://www.deri.ie/sites/default/files/publications/freitas_qald_2011_0.pdf">https://www.deri.ie/sites/default/files/publications/freitas_qald_2011_0.pdf</a></p> <p>Steve Harris and Andy Seaborne. Sparql 1.1 query language. Technical report, W3C, January 2012.  <a href="http://www.w3.org/TR/2012/PR-sparql11-query-20121108">http://www.w3.org/TR/2012/PR-sparql11-query-20121108</a></p>		
<b>Week 6 Sep 28</b>	<p><b>Large, Real-World KGs</b></p> <p><b>Project Proposal</b></p>	<p>S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. <a href="#">Dbpedia: A Nucleus for a Web of Open Data</a>. The Semantic Web, 2007.</p> <p>Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J. <a href="#">Freebase: a collaboratively created graph database for structuring human knowledge</a>. SIGMOD, 2008.</p> <p>Optional:</p> <p>X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmman, S. Sun, and W. Zhang. <a href="#">Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion</a>. KDD, 2014.</p> <p>F. M. Suchanek, G. Kasneci, and G. Weikum. <a href="#">Yago: A Core of Semantic Knowledge</a>. WWW, 2007.</p> <p>F. Niu, C. Zhang, C. Re, and J. W. Shavlik. <a href="#">DeepDive: Web-Scale Knowledge-Base Construction Using Statistical Learning and Inference</a>. VLDS, 2012.</p> <p>D. B. Lenat. <a href="#">Cyc: A Large-Scale Investment in Knowledge Infrastructure</a>. Communications of the ACM, 38(11):33–38, 1995</p>	<p>Quiz 5</p> <p>Project Proposals Due</p>	Pujara
<b>Week 7 Oct 5</b>	<b>Information Extraction</b>	<p>D. C. Wimalasuriya and D. Dou. Ontology-based Information Extraction: An Introduction and a Survey of Current Approaches. J. Information Science, 36(3), 2010.</p> <p>J. Zhu, Z. Nie, X. Liu, B. Zhang, and J.-R. Wen. Statsnowball: A Statistical Approach to Extracting Entity Relationships. In Proceedings of the 18th International Conference on World Wide Web, pages 101–</p>	<p>Quiz 6</p> <p>Homework 7: Information Extraction II</p>	Pujara



		<p>110. ACM, 2009.</p> <p>O. Etzioni, M. Banko, S. Soderland, and D. S. Weld. Open Information Extraction from the Web. <i>Communications of the ACM</i>, 51(12):68–74, 2008.</p> <p>S. Krause, H. Li, H. Uszkoreit, and F. Xu. Large-Scale Learning of Relation-Extraction Rules with Distant Supervision from the Web. <i>International Semantic Web Conference</i>, pages 263–278, 2012.</p> <p>AnHai Doan, Alon Y. Halevy, and Zachary G. Ives. <i>Principles of Data Integration</i>, chapters 4. Morgan Kaufmann, 2012.  <a href="http://www.sciencedirect.com/science/book/9780124160446">http://www.sciencedirect.com/science/book/9780124160446</a></p>		
<b>Week 8 Oct 12</b>	<b>Structured Data</b>	<p>AnHai Doan, Alon Y. Halevy, and Zachary G. Ives. <i>Principles of Data Integration</i>, chapters 7. Morgan Kaufmann, 2012.  <a href="http://www.sciencedirect.com/science/book/9780124160446">http://www.sciencedirect.com/science/book/9780124160446</a></p> <p>Limaye, G., Sarawagi, S., Chakrabarti, S.: Annotating and Searching Web Tables using Entities, Types and Relationships. <i>Proc. VLDB Endow.</i> 3(1-2), 1338-1347  <a href="http://www.vldb.org/pvldb/vldb2010/pvldb_vol3/R118.pdf">http://www.vldb.org/pvldb/vldb2010/pvldb_vol3/R118.pdf</a></p> <p>Cafarella, Michael J., Alon Halevy, Daisy Zhe Wang, Eugene Wu, and Yang Zhang. "Webtables: exploring the power of tables on the web." <i>Proceedings of the VLDB Endowment</i> 1, no. 1 (2008): 538-549.  <a href="http://tangra.si.umich.edu/~radev/767w10/papers/Week06/TextRepresentation/Cafarella.pdf">http://tangra.si.umich.edu/~radev/767w10/papers/Week06/TextRepresentation/Cafarella.pdf</a></p> <p>Crestan, Eric, and Patrick Pantel. "Web-scale table census and classification." In <i>Proceedings of the fourth ACM international conference on Web search and data mining</i>, pp. 545-554. ACM, 2011.  <a href="http://www.patrickpantel.com/download/papers/2011/wsdm11.pdf">http://www.patrickpantel.com/download/papers/2011/wsdm11.pdf</a></p>	<p>Quiz 7</p> <p>Homework 8: Structured Data</p>	Pujara

<p><b>Week 9</b> <b>Oct 19</b></p>	<p><b>Entity Resolution</b></p>	<p>I. P. Fellegi and A. B. Sunter. A theory for record linkage. Journal of the American Statistical Association, 64(328):1183–1210, 1969.</p> <p>W. E. Winkler. The state of record linkage and current research problems. In Statistical Research Division, US Census Bureau. Citeseer, 1999.</p> <p>G. Navarro. A Guided Tour to Approximate String Matching. ACM Comput. Surv., 33 (1):31–88, Mar. 2001. ISSN 0360-0300. doi: 10.1145/375360.375365. URL <a href="http://doi.acm.org/10.1145/375360.375365">http://doi.acm.org/10.1145/375360.375365</a></p> <p>Matthew Michelson and Craig A. Knoblock. Semantic Annotation of Unstructured and Ungrammatical Text. In Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-2005), Edinburgh, Scotland, 2005. <a href="http://www.isi.edu/integration/papers/michelson05-ijcai.pdf">http://www.isi.edu/integration/papers/michelson05-ijcai.pdf</a></p> <p>Andrew McCallum. Information Extraction: Distilling Structured Data from Unstructured Text . ACM Queue, volume 3, Number 9, November 2005. <a href="http://people.cs.umass.edu/~mccallum/papers/acm-queue-ie.pdf">http://people.cs.umass.edu/~mccallum/papers/acm-queue-ie.pdf</a></p> <p>Bo Wu, Pedro Szekely, and Craig A. Knoblock. Minimizing user effort in transforming data by example. In Proceedings of the International Conference on Intelligent User Interface, 2014. <a href="http://www.isi.edu/integration/papers/wu14-iui.pdf">http://www.isi.edu/integration/papers/wu14-iui.pdf</a>.</p> <p>Open Refine, Explore data. <a href="http://youtu.be/B70J_H_zAWM">http://youtu.be/B70J_H_zAWM</a>.</p> <p>Open Refine, Clean and transform data. <a href="http://youtu.be/cO8NVCs_Ba0">http://youtu.be/cO8NVCs_Ba0</a>.</p> <p>Open Refine, Reconcile and match data. <a href="http://youtu.be/5tsyz3ibYzk">http://youtu.be/5tsyz3ibYzk</a>.</p>	<p>Quiz 8</p> <p>Homework 9: Entity Resolution</p>	<p>Szekely</p>
<p><b>Week 10</b></p>	<p><b>Ontologies &amp;</b></p>	<p>Frank Manola and Eric Miller. Rdf primer.</p>	<p>Quiz 9</p>	<p>Szekely</p>

Oct 26	RDF	<p>Technical report, W3C, February 2004.  <a href="http://www.w3.org/TR/2004/REC-rdf-primer-20040210/">http://www.w3.org/TR/2004/REC-rdf-primer-20040210/</a></p> <p>Tim Berners-Lee. Why rdf model is different from the xml model. Technical report, W3C, 1998. <a href="http://www.w3.org/DesignIssues/RDF-XML.html">http://www.w3.org/DesignIssues/RDF-XML.html</a>.</p> <p>Rdf vocabulary description language 1.0: Rdf schema. Technical report, W3C, February 2004. <a href="http://www.w3.org/TR/2004/REC-rdf-schema-20040210/">http://www.w3.org/TR/2004/REC-rdf-schema-20040210/</a></p> <p>Ben Adida, Ivan Herman, Manu Sporny, and Mark Birbeck. Rdfa 1.1 primer rich structured data markup for web documents. Technical report, W3C, June 2012.  <a href="http://www.w3.org/TR/2012/NOTE-rdfa-primer-20120607/">http://www.w3.org/TR/2012/NOTE-rdfa-primer-20120607/</a></p>	Homework 10: Ontologies	
Week 11 Nov 2	Semantic Modeling	<p><a href="#">Pham, M.</a>; <a href="#">Alse, S.</a>; <a href="#">Knoblock, C.</a>; and <a href="#">Szekely, P.</a> Semantic labeling: A domain-independent approach. In <i>Proceedings of the 15th International Semantic Web Conference</i>, 2016.  <a href="http://usc-isi-i2.github.io/papers/pham16-iswc.pdf">http://usc-isi-i2.github.io/papers/pham16-iswc.pdf</a></p> <p>Mark James Carman and Craig A. Knoblock. Learning semantic descriptions of web information sources. In <i>Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)</i>, January 2007.  <a href="http://www.isi.edu/integration/papers/carm07-ijcai.pdf">http://www.isi.edu/integration/papers/carm07-ijcai.pdf</a></p> <p>José Luis Ambite, Sirish Darbha, Aman Goel, Craig A. Knoblock, Kristina Lerman, Rahul Parundekar, and Thomas Russ. Automatically constructing semantic web services from online sources. In <i>Proceedings of the 8th International Semantic Web Conference (ISWC 2009)</i>, 2009.  <a href="http://www.isi.edu/integration/papers/ambite09-iswc.pdf">http://www.isi.edu/integration/papers/ambite09-iswc.pdf</a></p> <p>Craig A. Knoblock and Pedro Szekely. Exploiting semantics for big data integration. <i>AI Magazine</i>, 2015.  <a href="http://usc-isi-i2.github.io/papers/knoblock15-">http://usc-isi-i2.github.io/papers/knoblock15-</a></p>	Quiz 10  Homework 11: Semantic Modeling	

		<p><a href="#">aimagazine.pdf</a></p> <p>Krtzsch Markus, Simancik Frantisek, and Horrocks Ian. A description logic primer. 2012. <a href="http://arxiv.org/pdf/1201.4089.pdf">http://arxiv.org/pdf/1201.4089.pdf</a>.</p> <p>R2rml: Rdb to rdf mapping language. <a href="http://www.w3.org/TR/r2rml/">http://www.w3.org/TR/r2rml/</a></p>		
<b>Week 12 Nov 9</b>	<b>Graph Embeddings &amp; Probabilistic Models</b>	<p>J. Pujara, H. Miao, L. Getoor, and W. Cohen. Using Semantics &amp; Statistics to Turn Data into Knowledge. AI Magazine, 36(1):65–74, 2015b</p> <p>Antoine Bordes, Nicolas Usunier, Alberto GarciaDuran, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In NIPS</p> <p>Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M Marlin. 2013. Relation extraction with matrix factorization and universal schemas. In NAACL-HLT.</p> <p>Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In NIPS</p> <p>Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In AAA</p> <p><a href="#">Random walk inference and learning in a large scale knowledge base</a> N Lao, T Mitchell, WW Cohen. EMNLP 2011 <a href="#">Incorporating vector space similarity in random walk inference over knowledge bases</a> M Gardner, P Talukdar, J Krishnamurthy, T Mitchell. EMNLP 2014.</p> <p>M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich. A Review of Relational Machine Learning for Knowledge Graphs: From Multi-Relational Link Prediction to Automated Knowledge Graph Construction.</p>	<p>Quiz 11</p> <p>Homework 12: Graph Embeddings</p>	Pujara

<b>Week 13 Nov 16</b>	<b>Search &amp; Queries</b>  <b>Intellectual Property</b>	<p>Rajaraman, J. Leskovec and J. D. Ullman, <i>Mining of Massive Datasets</i>, Cambridge University Press, 2012.  <a href="http://infolab.stanford.edu/~ullman/mmds/ch10.pdf">http://infolab.stanford.edu/~ullman/mmds/ch10.pdf</a></p> <p>Thomas P. Vartanian and Robert H. Ledig. Scrape it, scrub it and show it: The battle over data aggregation.  <a href="http://web.archive.org/web/20070818130311/http://www.ffhsj.com/bancmail/bmarts/abaart.html">http://web.archive.org/web/20070818130311/http://www.ffhsj.com/bancmail/bmarts/abaart.html</a>.</p> <p>Kembrew McLeod. Intellectual property law, freedom of expression, and the web, 2003.  <a href="http://www.electronicbookreview.com/thread/technocapitalism/proprietary">http://www.electronicbookreview.com/thread/technocapitalism/proprietary</a>.</p> <p>Electronic frontier foundation.  <a href="http://www.eff.org/issues/intellectual-property">http://www.eff.org/issues/intellectual-property</a>.</p>	Quiz 12	Szekely
<b>Week 14 Nov 23</b>	<b>Thanksgiving break! No class.</b>			
<b>Week 15 Nov 30</b>	<b>Student Presentations</b>			Szekely Pujara Vu
<b>FINAL Dec 7 2-4pm</b>	<b>Final Exam</b>		During assigned time in the <i>Schedule of Classes</i> at <a href="https://classes.usc.edu/term-20183/finals/">https://classes.usc.edu/term-20183/finals/</a>	

## Statement on Academic Conduct and Support Systems

### Academic Conduct

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Section 11, *Behavior Violating University Standards* <https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions>. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, <http://policy.usc.edu/scientific-misconduct>.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* <http://equity.usc.edu> or to the *Department of Public Safety* <http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us>. This is important for the safety of the whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Center for Women and Men* <http://www.usc.edu/student-affairs/cwm/> provides 24/7 confidential support, and the sexual assault resource center webpage <http://sarc.usc.edu> describes reporting options and other resources.

### Support Systems

A number of USC’s schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* <http://dornsife.usc.edu/ali>, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* [http://sait.usc.edu/academicsupport/centerprograms/dsp/home\\_index.html](http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html) provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information* <http://emergency.usc.edu> will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.