



Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Planilla a completar para presentación de Cursos de Posgrado

1.- DEPARTAMENTO de COMPUTACIÓN

2.- NOMBRE DEL CURSO: .. Tecnologías Semánticas para Respuesta Automática a Preguntas

3.- DOCENTES:

RESPONSABLE/S: .. Siddharth Patwardhan.....

COLABORADORES:

AUXILIARES:

4.- CARRERA de DOCTORADO

5.- AÑO: 2016.....

CUATRIMESTRE/S: Segundo.....

6.- PUNTAJE PROPUESTO PARA CARRERA DE DOCTORADO: $\frac{1}{2}$ punto

7.- DURACIÓN (anual, cuatrimestral, bimestral u otra): .1 Semana.....

8.- CARGA HORARIA SEMANAL:

Teóricas:

Problemas:

Laboratorio:

Seminarios:

Teórico – Práctico: ..15.....

Salida a Campo:

9.- CARGA HORARIA TOTAL:15hs.....

10.- FORMA DE EVALUACIÓN: ..Trabajo final

11.- PROGRAMA ANALÍTICO:

1. Introducción general a la arquitectura de un sistema automático de Respuesta a Preguntas (DeepQA). Vamos a cubrir brevemente algunos antecedentes y conocimientos

previos involucrados en la arquitectura de un sistema de DeepQA. Vamos a discutir las diversas decisiones de diseño realizados a lo largo del desarrollo del sistema Watson, y estudiar los pros y los contras de estas decisiones.

2. Componentes de un sistema automático de Respuesta a Preguntas. En esta sesión, vamos a describir los diferentes componentes que se ejecutan dentro del sistema de DeepQA. Vamos a discutir, en profundidad, las técnicas de Procesamiento de Lenguaje Natural (PLN) y algoritmos de aprendizaje automático que alimentan estos componentes, y también estudian el impacto de algunos de estos en un sistema automático de Respuesta a Preguntas.

3. La implementación del sistema con UIMA y su escalabilidad. Esta sesión cubrirá los detalles de implementación de un sistema automático de Respuesta a Preguntas. Vamos a seguir el proceso de construcción de tal sistema a partir de cero utilizando el framework de código abierto Apache UIMA. También vamos a discutir una estrategia para leveraging Apache UIMA-AS para escalar el sistema a través de 1000 de los núcleos.

4. Metodología para adaptar un sistema automático de Respuesta a Preguntas como Watson a otros dominios. En este punto se discutirá una metodología general para analizar y desarrollar tales sistemas.

5. Tópicos avanzados. Resolviendo cuestiones específicas y aplicaciones comerciales. Algunos tópicos versarán sobre técnicas de descomposición de preguntas y de manejo de casos especiales.

12.- BIBLIOGRAFÍA:

C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. DBpedia—Crystallization Point for the Web of Data. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165, September. 2009.

J. Chu-Carroll, J. Fan, B. Boguraev, D. Carmel, D. Sheinwald, and C. Welty. Finding Needles in the Haystack: Search and Candidate Generation. *IBM Journal of Research and Development*, 56(3.4):6:1–6:12. . 2012.

J. Chu-Carroll, J. Fan, N. Schlaefel, and W. Zadrozny. Textual Resource Acquisition and Engineering. *IBM Journal of Research and Development*, 56(3.4):4:1–4:11. 2012.

E. Epstein, M. Schor, B. Iyer, A. Lally, E. Brown, and J. Cwiklik. Making Watson Fast. *IBM Journal of Research and Development*, 56(3.4):15:1–15:12. 2012.

J. Fan, A. Kalyanpur, D. Gondek, and D. Ferrucci Automatic Knowledge Extraction from Documents. *IBM Journal of Research and Development*, 56(3.4):5:1–5:10. . 2012.

D. Ferrucci and A. Lally. UIMA: an Architectural Approach to Unstructured Information Processing in the Corporate Research Environment. *Natural Language Engineering*, 10(3-4):327–348. 2004.

D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, and C. Welty. Building Watson: An Overview of the DeepQA project. *AI magazine*, 31(3):59–79. 2010.

D. Ferrucci, A. Levas, S. Bagchi, D. Gondek, and E. Mueller. 2012. Watson: Beyond Jeopardy. *Artificial Intelligence* (in press). D. Ferrucci. Introduction to “This is Watson”. *IBM Journal of Research and Development*, 56(3.4):1:1–1:15. 2012.

A. Gliozzo and T. Isabella. *Semantic Domains in Computational Linguistics*. Technical report. 2005.

A. Kalyanpur, B. Boguraev, S. Patwardhan, J. W. Murdock, A. Lally, C. Welty, J. Prager, B. Coppola, A. Fokoue-Nkoutche, L. Zhang, Y. Pan, and Z. Qiu. Structured Data and Inference in DeepQA. *IBM Journal of Research and Development*, 56(3.4):10:1–10:14. 2012.

A. Lally, J. Prager, M. McCord, B. Boguraev, S. Patwardhan, J. Fan, P. Fodor, and J. Chu-Carroll. Question Analysis: How Watson Reads a Clue. *IBM Journal of Research and Development*, 56(3.4):2:1–2:14. 2012.

T. Landauer and S. Dumais. A Solution to Plato’s Problem: the Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge. *Psychological Review*, 104(2):211–240. 1997.

B. Lewis. In the Game: The Interface between Watson and Jeopardy! *IBM Journal of Research and Development*, 56(3.4):17:1–17:6. 2012.

M. McCord, J. W. Murdock, and B. Boguraev Deep Parsing in Watson. *IBM Journal of Research and Development*, 56(3.4):3:1–3:15. . 2012.

T. Miller, C. Biemann, T. Zesch, and I. Gurevych. Using Distributional Similarity for Lexical Expansion in Knowledge-based Word Sense Disambiguation. In *Proceedings of the International Conference on Computational Linguistics*, pages 1781–1796, Mumbai, India, December. 2012.

J. W. Murdock, J. Fan, A. Lally, H. Shima, and B. Boguraev. Textual Evidence Gathering and Analysis. *IBM Journal of Research and Development*, 56(3.4):8:1–8:14. 2012.

J. W. Murdock, A. Kalyanpur, C. Welty, J. Fan, D. Ferrucci, D. Gondek, L. Zhang, and H. Kanayama. Typing Candidate Answers Using Type Coercion. *IBM Journal of Research and Development*, 56(3.4):7:1–7:13. 2012.

J. Prager, E. Brown, and J. Chu-Carroll. Special Questions and Techniques. *IBM Journal of Research and Development*, 56(3.4):11:1–11:13. 2012.

G. Tesauro, D. Gondek, J. Lenchner, J. Fan, and J. Prager. 2012. Simulation, Learning, and Optimization Techniques in Watson’s Game Strategies. *IBM Journal of Research and Development*, 56(3.4):16:1–16:11. 2012.