



Universidade Federal de Santa Catarina  
Centro Tecnológico  
Departamento de Informática e Estatística  
Programa de Pós-Graduação em Ciência da Computação



## Plano de Ensino

### 1) Identificação

**Disciplina:** INE410118 - Tópicos Especiais em Computação: Semântica e Recuperação de Informação

**Carga horária:** 45 horas/aula – 3 créditos

**Professora:** Renato Fileto

2) **Requisitos:** não há.

### 3) Ementa:

Ontologias. Formalismos e padrões para representação e manipulação de conhecimento. Anotações livres e semânticas. Processamento de consultas baseado em semântica. Fundamentos da recuperação de informação. Modelos para recuperação de informação. Recuperação de dados complexos por similaridade de conteúdo. Espaços dimensionais. Espaços métricos. Processamento de consultas formuladas como composições de predicados heterogêneos.

### 4) Objetivos:

**Geral:** Capacitar o aluno em conceitos e técnicas para manipulação de semântica de dados e para recuperação de informação, desenvolvendo habilidades em tecnologias afins e senso crítico para conduzir pesquisas envolvendo uma dessas áreas ou a intersecção delas.

### Específicos:

Ao final do curso, espera-se que o aluno seja capaz de:

1. Conhecer os principais conceitos e técnicas relacionados com semântica de dados;
2. Dominar linguagens para representação e manipulação de conhecimento segundo formalismos da lógica descritiva;
3. Conhecer os principais conceitos, modelos e técnicas utilizados na recuperação de informação;
4. Dominar técnicas e ferramentas básicas para estipulação e processamento de consultas em sistemas de recuperação de informação.
5. Compreender o estado da arte das pesquisas em manipulação de semântica e recuperação de informação.

### 5) Conteúdo Programático:

- Conceitos fundamentais relacionados com a semântica de dados [3 horas-aula].
- Formalismos e padrões para a representação do conhecimento [3 horas-aula].

- Técnicas e ferramentas baseadas em lógica descritiva para explicitação e manipulação de conhecimento [6 horas-aula].
- Panorama das pesquisas em manipulação de semântica [6 horas-aula].
- Conceitos, modelos e técnicas utilizados na recuperação de informação [6 horas-aula].
- Buscas semânticas [6 horas-aula].
- Panorama das pesquisas recuperação de informação de grandes volumes de dados complexos [6 horas-aula].
- Orientação, acompanhamento e apresentação de trabalhos dos alunos [9].

## 6) Metodologia:

As aulas serão expositivas, estimulando o diálogo e interações entre alunos e o professor, para que o entendimento e a crítica apareçam como resultado de aprendizagem. Os estudos partirão de problemas do mundo real, em uma abordagem construtivista, sempre que possível. Além de slides, exemplos, software e outros materiais da coleção do professor, serão usados recursos didáticos e de software desenvolvidos colaborativamente e disponíveis na forma de *OpenCourseWare*, tais como:

- Information Retrieval - <http://slidewiki.org/deck/345#tree-0-deck-345-1-view>
- Semantic Data Web - <http://slidewiki.org/deck/750#tree-0-deck-750-1-view>

Serão propostos exercícios em sala de aula, em laboratório e principalmente extra-classe, como forma de fixar os conceitos apresentados.

A ferramenta Protégé será utilizada para a edição de bases de conhecimento em RDF e OWL. Posteriormente, serão utilizados repositórios de triplas, como o Parliament, o Jena, o Sesame, para a integração das bases de conhecimento produzidas com conjuntos de dados ligados (linked data) disponíveis na Web, tais como DBPedia, GeoNames e LinkedGeoData. Será utilizada a linguagem SPARQL para a realização de consultas sobre a base de conhecimento e a dados ligados. Também serão utilizadas ferramentas para visualização de conhecimento e cálculo de métricas de similaridade semântica sobre ontologias disponíveis na Web como a WordNet. Outras ferramentas, máquinas de inferência, ferramentas com extensões para o tratamento de dados geográficos ou similaridade poderão ser introduzidas de acordo com o interesse dos alunos.

Para cumprir o item 5 dos objetivos específicos, os alunos farão pesquisa bibliográfica, leitura de artigos e confecção de resenha ou proposta de pesquisa, devidamente formal e embasada na literatura técnico-científica. A definição dos possíveis grupos de alunos e a definição do tema específico de cada aluno ou grupo se fará em colaboração com o professor, dentro de prazos estabelecidos na primeira semana de aulas. O acompanhamento dos trabalhos será feito de acordo com um cronograma estabelecido no início do curso, prevendo as datas para revisões das tarefas e dos diferentes estágios do desenvolvimento do trabalho e das apresentações finais.

Sendo esta uma disciplina de pós-graduação, espera-se do aluno um comportamento proativo, com dedicação em tempo igual ou maior que o número de horas-aulas da disciplina para estudo e pesquisa do conteúdo apresentado em sala de aula.

## 7) Avaliação:

As seguintes **tarefas** serão **avaliadas**:

1. Cada aluno deve selecionar, resumir e avaliar as contribuições de 1 artigo no tema específico de seu estudo, de acordo com os critérios definidos pelo professor no início das aulas. Cada

aluno deve entregar seu resumo e avaliação de artigo juntamente com cópia dos artigo selecionado.

2. Apresentação de até 30 minutos de duração sobre o tema estudado, incluindo definição do problema abordado, pesquisa bibliográfica com quadro comparativo das abordagens propostas na literatura e descrição da abordagem adotada no trabalho.
3. Redação de artigo científico no tema estudado. O artigo deve ser objetivo, claro, organizado, ser bem redigido e ter rigor científico. Deve explicar concisamente o problema tratado e a abordagem adotada desde o resumo e introdução. Também deve enumerar suas contribuições e trabalhos futuros nas conclusões.

Apuração do rendimento dos alunos será feita de acordo com a seguinte fórmula:

$$\text{Nota final} = 0,3 * \text{tarefa1} + 0,3 * \text{tarefa2} + 0,4 * \text{tarefa3}$$

## 8) Cronograma

- 1ª semana: apresentação da disciplina e introdução à semântica de dados;
- 2ª semana: introdução a RDF, OWL e SPARQL;
- 3ª semana: prática sobre o Protegé;
- 4ª semana: prática sobre o Protegé e repositório RDF;
- 5ª semana: pesquisas sobre manipulação de semântica e definição da leitura;
- 6ª semana: breves apresentações (15 min.cada) e discussão de resumos de leituras;
- 7ª semana: conceitos e modelos de recuperação de informação;
- 8ª semana: estudos de modelos e técnicas específicos de recuperação de informação;
- 9ª semana: conceitos e técnicas fundamentais de buscas semânticas;
- 10ª a 13ª semana: pesquisas em técnicas e ferramentas de buscas semânticas e recuperação de informação de grandes bases de dados complexos;
- 14ª a 16ª semana: apresentações dos trabalhos;
- 17ª a 18ª semana: revisão da redação dos artigos.

## 9) Bibliografia:

- **Bibliografia Básica**
  - Antoniou, G., and van Harmelen, F. (2012). A Semantic Web Primer, 3rd edition. The MIT Press, Cambridge, MA, USA. (<http://www.springerlink.com/content/978-3-540-76451-9>)
  - Hitzler, Krötzsch, Rudolph, (2009). Foundations of Semantic Web Technologies, Chapman & Hall/CRC. (<http://www.semantic-web-book.org>)
  - Baeza-Yates, R. A. and Ribeiro-Neto, B. A. (2011). Modern Information Retrieval - the concepts and technology behind search, 2nd edition. Pearson Education Ltd., Harlow, England. (<http://people.ischool.berkeley.edu/~hearst/irbook>)
  - Manning, C. D. , Raghavan P. and Schütze H. (2008). Introduction to Information Retrieval, Cambridge University Press. (<http://nlp.stanford.edu/IR-book>)
- **Bibliografia Complementar**
  - Baeza-Yates, R. and Melucci, M. (Eds.) (2011). Advanced Topics in Information Retrieval. Springer.
  - Blanken, H., de Vries, A., Blok, H., and Feng, L., editors (2007). Multimedia Retrieval.
  - Data-Centric Systems and Applications. Springer Verlag, Heidelberg.

- Breitman, K.K., Casanova, M.A., Truskowski, W. (2007). Semantic Web: Concepts, Technologies and Applications. Series: NASA Monographs in Systems and Software Engineering, Springer.
- Davies, J., Studer, R., Warren, P. (Eds.) (2006). Semantic Web Technologies: trends and research in ontology-based Systems, John Wiley & Sons.
- Hitzler, P., Krotzsch, M., Rudolph, S. (2009). Foundations of Semantic Web Technologies. Chapman & Hall/CRC.
- Kashyap, V., Bussler, C., Moran, M. (2008). The Semantic Web - Semantics for Data and Services on the Web. Series: Data-Centric Systems and Applications. Springer.
- Zezula, P., Amato, G., Dohnal, V., and Batko, M. (2006). Similarity Search - The Metric Space Approach, volume 32. Springer.