UNIVERSIDADE FEDERAL DE SANTA CATARINA – UFSC
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO – CIN

CIN 7601 – PLANO DE ENSINO (2021.2)

1 IDENTIFICAÇÃO

DISCIPLINA: CIN 7601 – Linked Data
CARGA HORÁRIA: 36h semestrais / 2h semanais
PROFESSOR: Moisés Lima Dutra (*moises.dutra@ufsc.br*)
OFERTA: Obrigatória para a 6ª. Fase do Curso de Ciência da Informação
HORA E LOCAL: Oferta remota em 2021.2, com atividades síncronas e assíncronas
As atividades síncronas ocorrerão nas quartas-feiras, das 08h20 às 10h00

2 EMENTA

Práticas de *Linked Data*. Criação, compartilhamento e utilização de dados e informações semanticamente conectadas. *Datasets* e *triplestores*. *Linked Open Data*.

3 OBJETIVOS

3.1 Objetivo Geral
Desenvolver a capacidade de criar, publicar, compartilhar e utilizar dados ligados, dentro dos princípios estabelecidos e aceitos para tal.

3.2 Objetivos Específicos
3.2.1 Conhecer os princípios dos dados ligados.
3.2.2 Estudar o modelo de dados RDF, visando a estruturação de *datasets*.
3.2.3 Exercitar a construção de *triplestores*, a partir da utilização de vocabulários abertos.
3.2.4 Aprender a utilizar a linguagem de busca semântica SPARQL.
3.2.5 Compreender as etapas de publicação e utilização de dados ligados abertos.

4 CONTEÚDO PROGRAMÁTICO

4.1 Princípios do *Linked Data*
4.1.1 Contextos de Utilização
4.1.2 URIs e IRIs
4.1.3 Espaços de Nomes
4.1.4 Topologia de Dados
4.1.5 Vocabulários para a Descrição de Dados

4.2 RDF (*Resource Description Framework*)
4.2.1 Modelo de Dados RDF
4.2.2 Formatos RDF para *Linked Data*
4.2.3 Definição de *Datasets*
4.2.4 Construção de *Triplestores* (bases de dados RDF)

4.3 SPARQL (*SPARQL Protocol and RDF Query Language*)
4.3.1 APIs SPARQL
4.3.2 Sintaxe SPARQL
4.3.3 SPARQL *Endpoints*

4.4 Linked Open Data
4.4.1 Dados Abertos
4.4.2 Vocabulários Abertos
4.4.3 LOD *Cloud*
4.4.4 *Data Hub*
4.4.5 Dados Abertos Governamentais

5 METODOLOGIA

Aulas expositivas e resolução de atividades práticas, baseadas na criação de *datasets* e na consulta destes.

As atividades remotas consistirão de quatro tipos:

- Aulas expositivas (**atividade síncrona**): 1 ou 2 horas-aula
- Resolução de atividades práticas e listas de exercícios (**atividade assíncrona**): 1 ou mais semanas
- Retiradas de dúvidas e acompanhamento do desenvolvimento das atividades (**atividade síncrona**): 1 ou 2 horas-aula
- Provas individuais (**atividade assíncrona**): 24 horas

Ambiente Virtual de Aprendizagem (principal instrumento de suporte da disciplina):

- **Moodle** (http://moodle.ufsc.br/). Deve ser **consultado todas as semanas** pelos alunos.

Ferramentas de suporte para as aulas remotas:

- Moodle (ferramenta-base)
- Google Meet (primordialmente, para as atividades síncronas)
- Chat RNP (eventualmente)
- Youtube (eventualmente)

Ferramentas on-line para o desenvolvimento das atividades práticas:

- Sublime Text (sugerida, gratuita e necessita instalação: https://www.sublimetext.com/)
- GraphDB Free (sugerida, gratuita e necessita instalação: https://graphdb.ontotext.com/)
- Google Colab (sugerida, acesso on-line em: https://colab.research.google.com/)
- Outra (sujeita a aprovação pelo professor)

Importante: **Não haverá gravação de conteúdo das aulas remotas**! Pede-se aos estudantes que **observem e respeitem** esta diretiva!

6 AVALIAÇÃO

Frequência: **Cada estudante será responsável por marcar suas presenças no Moodle da disciplina** (esta opção estará disponível por 24 horas **no dia exato de cada aula**). Se programem para fazer isso!

Conforme a Resolução do Conselho Universitário 017/Cun/97, de 30/09/97:

- ❖ A frequência mínima obrigatória é de 75% das aulas. Enquanto vigorar o Calendário Suplementar Excepcional da UFSC, a presença física na aula será comutada pela autoanotação de presença do estudante no Moodle.

Avaliações:

- Serão realizadas **duas avaliações individuais assíncronas**: Prova 1 (**P1**) e Prova 2 (**P2**).
- **As questões serão disponibilizadas no Moodle no dia marcado no cronograma e cada estudante terá um prazo de 24 horas para enviar as respostas via Moodle**.

A média final (**MF**) será calculada conforme a seguinte fórmula:

$$MF = 0.4(P1) + 0.6(P2)$$

- ❖ O aluno que obtiver a **MF entre 3,0 e 5,5** e frequência suficiente poderá, ao final do semestre, realizar uma prova de recuperação (avaliação individual assíncrona) de todo o conteúdo.
- ❖ **EM CASO DE RECUPERAÇÃO**, a nota final do semesre (**NFS**) será calculada a partir da **média simples** entre a média final (**MF**) obtida durante o semestre e a nota obtida na prova de recuperação (**PR**).

$$NFS = (MF + PR) / 2$$

7 BIBLIOGRAFIA

Básica:

AUER, Sören et al. **Introduction to Linked Data and its Lifecycle on the Web**. Disponível em: <http://jens-lehmann.org/files/2013/reasoning_web_linked_data.pdf>. Acesso em: 05 ago. 2020.

BAUER, Florian; KALTENBÖCK, Martin. **Linked Open Data: The Essentials**: A Quick Start Guide for Decision Makers. Disponível em: <http://www.reeep.org/LOD-the-Essentials.pdf>. Acesso em: 05 ago. 2020.

BERNERS-LEE, T. **Linked data - Design issues**. 2006, Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>, Acesso em: 05 ago. 2020.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. **The Semantic Web**: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American Special, número especial, 2001. Disponível em: <http://csis.pace.edu/~marchese/CS835/Lec9/112_SemWeb.pdf>. Acesso em: 05 ago. 2020.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. **Linked Data - The Story So Far**. Disponível em: <https://eprints.soton.ac.uk/271285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: 05 ago. 2020.

DBPEDIA TEAM. **DBPedia**. 2017. Disponível em <http://wiki.dbpedia.org/>. Acesso em: 05 ago. 2020.

HEATH, Tom; BIZER, Cristian. **Linked Data**: Evolving the Web into a Global Workspace. Disponível em: <https://pdfs.semanticscholar.org/da63/084ec46c1912c4b4c4cad1ec8104dbef74aa.pdf>. Acesso em: 05 ago. 2020.

LINKEDDATATOOLS.COM. **Free tools, information and resource for the semantic web**. Disponível em: <http://www.linkeddatatools.com>. Acesso em: 05 ago. 2020.

LOV. **Linked Open Vocabularies**. 2017. Disponível em: <http://lov.okfn.org/dataset/lov/>. Acesso em: 05 ago. 2020.

RDF. **Resource Description Framework (RDF)**, 2014. Disponível em: <http://www.w3.org/RDF/>. Acesso em: 05 ago. 2020.

SPARQL. **Sparql 1.1 Overview**. W3C, 2013. Disponível em <http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>. Acesso em: 05 ago. 2020.

TURTLE. **RDF 1.1 Turtle**, 2014. Disponível em: <https://www.w3.org/TR/turtle/>. Acesso em: 05 ago. 2020.

Complementar:

ALLEMANG, D.; HENDLER, J. **Semantic Web for the Working Ontologist**: Effective Modeling in RDFS and OWL. 2a. edição. Burlington (USA): Morgan Kaufmann, 2011. ISBN: 978-0123859655.

ANTONIOU, G. et al. **A Semantic Web Primer**: Cooperative Information Systems. 3a. edição. Cambridge (MA, USA): The MIT Press, 2012. ISBN: 978-0262018289.

APACHE. **Apache Jena**. 2017. Disponível em <https://jena.apache.org/>. Acesso em: 22 set. 2017.

AUER, Sören et al**. DBpedia: A Nucleus for a Web of Open Data**. Disponível em: <http://158.130.69.163/~zives/research/dbpedia.pdf>. Acesso em: 22 set. 2017.

CAFARELLA, M. J.; HALEVY, A.; MADHAVAN, J. Structured Data on the Web. **Communications of the ACM**, vol. 54, no. 2, fevereiro 2011.

DUCHARME, B. **Learning SPARQL**. Sebastopol (USA): O'Reilly Media, 2013. ISBN: 978-1449371432.

HEATH, Tom; BIZER, Christian. **Linked data**: envolving the web into a global data space. San Rafael, Calif.: Morgan & Calypool, c2011. xi, 122 p. (Synthesis lectures on the semantic web : theory and technology). ISBN 9781608454303.

HOOLAND S.V.; VERBORGH, R. **Linked Data for Libraries, Archives and Museums**: How to Clean, Link and Publish your Metadata. Chicago (USA): American Library Association Editions, 2014. ISBN: 978-0838912515.

N3. **RDF NOTATION3 (N3)**, 2011. Disponível em: <https://www.w3.org/TeamSubmission/n3/>. Acesso em: 22 set. 2017.

NOY, N.; MCGUINNESS, D. **Ontology Development 101: A Guide to Creating Your First Ontology**. Serra Mall, California, USA: Stanford University, March 2001. (Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880). Disponível em: <https://protege.stanford.edu/publications/ontology_development/ontology101.pdf>. Acesso em 22 set. 2017.

PROTÉGÉ TEAM. **Protégé: A free, open-source ontology editor and framework for building intelligent systems**. 2014. Disponível em: <http://protege.stanford.edu/>. Acesso em: 22 set. 2017.

W3C, 2013. **Data Activity**. W3C, 2017. Disponível em: <http://www.w3.org/2013/data/>. Acesso em: 22 set. 2017.

W3C, 2017a. **Linked Data**. W3C, 2015. Disponível em: <https://www.w3.org/standards/semanticweb/data>. Acesso em: 22 set. 2017.

W3C, 2017b. **Semantic Web**. W3C, 2015. Disponível em: <http://www.w3.org/standards/semanticweb/>. Acesso em: 22 set. 2017.

W3C, 2017c. **XML Technology**. W3C, 2015. Disponível em: <http://www.w3.org/standards/xml/>. Acesso em: 22 set. 2017.

W3C INCUBATOR GROUP. **Library Linked Data Incubator Group Final Report**. Cambridge, Ma: W3c, 2011. Disponível em: <https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>. Acesso em: 22 set. 2017.

WOOD, D. et al. **Linked Data**. Greenwich (USA): Manning Publications, 2014. ISBN: 978-1617290398.

8 CRONOGRAMA *		
AULA	**DATA**	**CONTEÚDO**
1	27/10	Aula Expositiva: Apresentação da Disciplina (**atividade síncrona**): 30 minutos. Aula Expositiva: Triplas e *Triplestores* (**atividade síncrona**): 1,5 hora-aula. Atividades Práticas (**atividade assíncrona**): 1 semana.
2	03/11	Aula Expositiva: Triplas e *Triplestores* (**atividade síncrona**): 1 hora-aula. Atividades Práticas (**atividade assíncrona**): 1 semana.
3	10/11	Aula Expositiva: Triplas e *Triplestores* (**atividade síncrona**): 1 hora-aula. Atividades Práticas (**atividade assíncrona**): 1 semana.
4	17/11	Aula Expositiva: *Triplas e Triplestores* (**atividade síncrona**): 1 hora-aula. Atividades Práticas (**atividade assíncrona**): 1 semana.
5	24/11	Aula Expositiva: Consultas Básicas (**atividade síncrona**): 1 hora-aula. Atividades Práticas (**atividade assíncrona**): 1 semana.
6	01/12	Aula Expositiva: Consultas Básicas (**atividade síncrona**): 1 hora-aula. Atividades Práticas (**atividade assíncrona**): 1 semana.
7	08/12	Aula Expositiva: Consultas Básicas (**atividade síncrona**): 1 hora-aula. Atividades Práticas (**atividade assíncrona**): 1 semana.
8	15/12	Aula de Acompanhamento: Retirada de Dúvidas (**atividade síncrona**): 2 horas-aula. Atividades Práticas (**atividade assíncrona**): 7 semanas.
RECESSO ESCOLAR		
9	02/02	**(P1): Prova 1** – via Moodle (**atividade assíncrona**): 24 horas.
10	09/02	Aula Expositiva: Linked Open Data (**atividade síncrona**): 1 hora-aula. Atividades Práticas (**atividade assíncrona**): 1 semana.
11	16/02	Aula Expositiva: Consultas Avançadas (**atividade síncrona**): 1 hora-aula. Atividades Práticas (**atividade assíncrona**): 3 semanas.
12	23/02	Estudo Prático Individual (**atividade assíncrona**).
13	02/03	**Feriado (Cinzas)**
14	09/03	**(P2): Prova 2** – via Moodle (**atividade assíncrona**): 24 horas.
15	16/03	**(PR): Prova de Recuperação** – via Moodle (**atividade assíncrona**): 24 horas.
16	23/03	**Feriado (Aniversário de Florianópolis)**

Cronograma sujeito a alterações e adaptações ao longo do semestre.

IMPORTANTE: Devido à excepcionalidade do semestre, poderão ser disponibilizados materiais de conteúdo e atividades via Moodle fora dos dias letivos ou durante as semanas que possuem feriado, de maneira a serem trabalhados de maneira assíncrona. Isso também ocorrerá quando alguma eventualidade impossibilitar a ocorrência da atividade síncrona previamente agendada!