



CENTRO UNIVERSITÁRIO CTC DEPARTAMENTO DE INFORMÁTICA		Ano e semestre 2018.1
INF2891	Tópicos de Hipertexto e Multimídia II Título do Tópico: Construção de “Knowledge Graphs” (Grafos de Conhecimento)	
	CARGA HORÁRIA TOTAL: 45	CRÉDITOS: 3
	Prof. Daniel Schwabe	

<b>OBJETIVOS</b>	Adquirir os conhecimentos necessários para projetar e construir base de dados conectados (“Linked Data”) semi-estruturados que formam grandes Grafos de Conhecimento (“Knowledge Graphs – KG”). Estes dados são extraídos de diversas fontes, incluindo documentos textuais, páginas web, dados em base de dados relacionais e NoSQL.
<b>EMENTA</b>	Os últimos anos evidenciam um esforço crescente de adquirir e integrar dados que são gerados massivamente na Web e na Internet, e utilizá-lo para alavancar diversos tipos de aplicações, tais como máquinas de busca (p.ex. Google, Bing, Yahoo, Yandex), redes sociais (p.ex., Facebook, LinkedIn, Google+, Instagram), bases de dados científicas, informações estatísticas, etc... Este curso abordará as técnicas para o projeto e implementação destas bases, conhecidas como Grafos de Conhecimento (“Knowledge Graphs”), que podem ter escopo tanto global quanto local a uma instituição ou a grupos de instituições. Os fundamentos e técnicas a serem abordados incluem raspagem de dados (“web crawling”) e extração de informações de páginas Web; extração de dados de documentos textuais; alinhamento de fontes de dados (semi) estruturados; ligação de entidades; base de dados em Grafo; limpeza e uniformização de dados; Web Semântica e Dados Conectados (“Linked Data”); consulta e navegação em Base de Dados em Grafos.
<b>PROGRAMA</b>	Introdução, exemplos Rastreamento (crawling) e raspagem de dados Extração de Informações Limpeza e homogeneização dos dados Modelagem Semântica, Grafos, RDF e RDFS, JSON-LD Mapeamento de modelos Similaridade e casamento de strings Ligação de Entidades Linked Data Proveniência Visão geral de consultas e análise de KGs Aspectos éticos e legais

<b>AVALIAÇÃO</b>	A avaliação será feita a partir das apresentações sobre tópicos relevantes feitas pelos participantes. Haverá também um trabalho final, que será um projeto de implementação a ser realizado em grupo, utilizando a infra estrutura do Laboratório TecWeb. Este projeto será integrado a um projeto já em andamento de construção de uma base pública de dados ligados acerca de agentes políticos no Brasil.
<b>BIBLIOGRAFIA PRINCIPAL</b>	AnHai Doan, Alon Y. Halevy, and Zachary G. Ives. Principles of Data Integration, chapters 7. Morgan Kaufmann, 2012 . <a href="http://www.sciencedirect.com/science/book/9780124160446">http://www.sciencedirect.com/science/book/9780124160446</a>  Linked Data – Evolving de Web into a Global Data Space – Tom Heath, Christian Bizer. Morgan&Claypool, 2011. <a href="http://euclid-project.eu/">http://euclid-project.eu/</a>
<b>BIBLIOGRAFIA COMPLEMENTAR</b>	Referencias coletadas e disponibilizadas no website do curso.