

Ontologia Aplicada à Universidade de Brasília - UnB Gama

Ontologia Aplicada à Universidade de Brasília - UnB Gama

RAFAEL FAZZOLINO PINTO BARBOSA - FAZZOLINO29@GMAIL.COM - 11/0136942
VICTOR HUGO ARNAUD DEON - VICTORHUGO@HOTMAIL.COM - 13/0136484
MURILO DUARTE - MURILODUARTEGONCALVES@HOTMAIL.COM - 12/0130645
ANTÔNIO CARVALHO DE OLIVEIRA - ANTONIOJK7@GMAIL.COM - 11/0058496
GABRIEL DOS SANTOS SILVA - GABRIEL93.SILVA@GMAIL.COM - 12/0011727

Sumário

1	Equipe	1
2	Apresentação do Documento do Projeto	1
3	Introdução	1
	3 .1 Fundamentação Teórica	1
4	Descrição do Problema	2
5	Objetivos do Projeto	2
	5 .1 Objetivos Gerais	2
	5 .2 Objetivos Específicos	3
6	Metodologia	3
	6 .1 Organização Interna	3
	6 .2 Descrição da Solução	3
	6 .3 Desenvolvimento da Ontologia	3
7	Descrição dos Usuários a Serem Beneficiados	4
8	Descrição do Ambiente Web	5
9	Cronograma do Projeto	5

1 Equipe

Os integrantes do projeto estão dispostos na tabela a seguir.

Nome	Email	Matrícula
Rafael Fazzolino	fazzolino29@gmail.com	11/0136942
Murilo Duarte Gonçalves	muriloduartegoncalves@gmail.com	12/0130645
Gabriel dos Santos Silva	gabriel93.silva@gmail.com	12/0011727
Antônio carvalho de oliveira Júnior	antoniojk7@gmail.com	11/0058496
Victor Hugo Arnaud Deon	victorhad@gmail.com	13/0136484

Tabela 1. Integrantes do Projeto

2 Apresentação do Documento do Projeto

Este documento tem como objetivo apresentar uma solução ao problema específico de Acompanhamento de Ementas Curriculares na Universidade de Brasília - UnB Gama. Contextualizando o Problema utilizando embasamento teórico referente ao desenvolvimento de uma Web Semântica no contexto de Ensino Superior de Qualidade.

O planejamento do desenvolvimento de uma Ontologia faz parte deste projeto, englobando desde a formulação das Necessidades até a obtenção de toda a especificação da Ontologia requerida.

3 Introdução

O desenvolvimento de conteúdo Web explodiu no mundo por volta da década de 90, onde todos os conteúdos eram desenvolvidos por programadores de *software* com objetivo de informar leitores Humanos. Desta forma, mecanismos de busca como *Google*, *AltaVista* entre outros, ainda necessitam da intervenção humana para que se possa identificar as respostas que realmente atendem as nossas demandas. A internet atual é considerada como Web Sintática, ou *web 2.0* e estamos caminhando para a web 3.0, pois nela os computadores fazem apenas a apresentação da informação a interpretação fica sob responsabilidade dos seres humanos.

O fato dos computadores não conseguirem interpretar essas informações se refere ao fato de que as páginas da web não têm informações sobre si mesmas, ou seja, que tipo de conteúdo está descrito e as informações sobre os assuntos na qual a página se refere.

Um bom exemplo do que estamos vivenciando hoje, neste contexto são os mecanismos de busca da *Google*, entre outros, que contém um grande número de páginas encontradas, porém com uma precisão muito pequena. Para melhorar isso é necessário adicionar semântica a essas páginas.

A utilização de Semântica para o desenvolvimento de sistemas web torna possível a interpretação do conteúdo por parte do computador, facilitando a busca de conteúdo que agregue, realmente, valor para o usuário.

3.1 Fundamentação Teórica

A Web Semântica é uma extensão da *World Wide Web* atual, que permitirá aos computadores e humanos trabalharem em cooperação, a web 1.0 possui como principal preocupação, a própria construção de rede, torná-la acessível e comercializável, criação de sites. Já a Web 2.0 é mais focada na colaboração *online* e na partilha entre utilizadores, rede sociais, blogs e etc, a 3.0 é que consideramos hoje de web inteligente, é uma web que se baseia na capacidade de interpretar conteúdos em rede, trazendo resultados mais precisos e inteligentes, mais conhecida como web Semântica.

Em um artigo de Tim Berners Lee [1], James Hendler e Ora Lassila, eles apresentaram cenários futuros onde a web semântica possui um papel fundamental em facilitar tarefas do cotidiano das pessoas. A ideia central da web semântica é categorizar a informação de maneira padronizada, facilitando seu acesso, é uma ideia bem semelhante à classificação dos seres vivos, essa web deve ser o mais descentralizada possível, a previsão de Hendler é de que qualquer empresa, universidade ou organização na web do futuro terá seu modelo próprio de organização, tendo assim uma série de modelos de organização em paralelo.

Um dos termos que tem grande importância na web semântica são os *metadados*, que são dados sobre dados, eles servem para indexar páginas e sites na web semântica, permitindo que outros computadores saibam de que assunto eles tratam, são informações sobre um certo dado como data, autor ou editora, que podem ser entendidos por máquinas.

Os metadados podem ser de 5 tipos, administrativo, descritivo, preservação, técnica e utilização, podemos também citar algumas características dos metadados, eles não precisam ser necessariamente digitais, podem ser

obtidos a partir de uma variedade de fontes, evoluem durante a vida útil do sistema de informação ou objeto a que se refere e vão além de fornecer dados sobre um objeto.

Temos vários formatos de captura de metadados, entre eles, *Dublin Core* que é composto por vários elementos como, assunto, título, criador, descrição do conteúdo do objeto, editor, entre outros, é um padrão bem simples, porém não apresenta uma semântica tão expressiva, outro formato de captura de metadados é o *Framework de Warwick* que surgiu pela necessidade de ampliar o *Dublin Core*, essa estrutura é baseada no conceito de contêineres, que agrega vários tipos de metadados em pacotes separados, o *Dublin Core* é apenas um desses pacotes, porém esse formato também tinha algumas falhas, então foi criado outro formato chamado RDF (*Resource Description Framework*), o RDF é uma linguagem declarativa que fornece uma maneira padronizada de utilizar o XML para representar metadados no formato de sentenças sobre propriedades e relacionamentos entre itens na web. O RDF foi projetado de modo a representar metadados de recursos web de maneira legível e, sobretudo, processável por máquinas.

Jim Hendler acredita que no futuro cada site e aplicação na internet vão contar com sua própria ontologia de termos, construídas e mantidas por pessoas, entidades ou instituições independentes e no futuro, os serviços providos na internet como reservas em hotéis, compras e etc, poderão ser grandemente expandidas e melhoradas se for adicionado semântica aos presentes recursos, as pessoas passarão a utilizar agentes, programas de software autônomos que agem em benefício de seus usuários, para realizar suas tarefas, porém quem irá tomar as decisões ainda somos nós, o seu papel será de reunir, organizar, selecionar e apresentar informações a um usuário humano, que tomará suas decisões, esses agentes farão uso das ontologias e metadados para realizar esses fins.

Vale ressaltar que web semântica não é uma Inteligência Artificial, não é uma web separada e não vai exigir que todas as aplicações utilizem expressões complexas. A web semântica vai ter um impacto gigantesco no mundo já que vai beneficiar tanto o comércio eletrônico como os usuários que poderão ter mais tempo livres já que a essa web irá trabalhar por ele.

4 Descrição do Problema

As Matrizes Curriculares apresentam quais disciplinas o aluno precisa cumprir para a conclusão de determinado curso. Cada unidade curricular possui uma ementa constituída pelos objetivos da disciplina, perfil dos ingressantes (pré-requisitos), e bibliografia básica e complementar.

O fluxo curricular e as ementas das disciplinas são de fundamental importância no processo de matrícula do aluno. No entanto, no contexto da Universidade de Brasília (UnB), há uma deficiência na apresentação dessas informações, resultando em dificuldades na obtenção de informações referentes às disciplinas e de relações entre as elas. No contexto específico à Faculdade do Gama (FGA), existe grande mutabilidade em relação às matrizes curriculares. A constante alteração no fluxo é um agravante para a pobre representação de informações.

Tais fatores levam a dificuldades no processo de matrícula. Além da confusão no controle do fluxo curricular, o aluno não possui auxílio ou base de comparação para a escolha de disciplinas optativas, o que pode levar à matrícula em disciplinas que não se encaixam em seu perfil. Falta de interesse e baixo rendimento são só algumas das consequências.

O problema se enquadra na ausência de uma representação semântica das informações Curriculares dos cursos de Engenharia da Universidade de Brasília - UnB Gama e de todas as Universidades brasileiras. Todo o trabalho de oferta e análise se baseia em nomes elencados e fluxos estáticos, não havendo nenhum sentido ou significado envolvidos. A abrangência deste problema vai deste ensino público até privado, englobando, todos os cursos e universidades que não possuem representação semântica de seus Currículos. O foco deste trabalho será no estudo do problema relacionado exclusivamente a Universidade de Brasília - Campus Gama, porém este estudo gera um *input* muito importante para a solução de diversos outros contextos.

5 Objetivos do Projeto

5.1 Objetivos Gerais

Transformação da aprendizagem em um processo autônomo e dinâmico, permitindo que os alunos construam seu histórico de acordo com seus objetivos, direcionando sua formação específica. Para isso é necessário que os estudantes tenham o máximo de informações possíveis acerca dos conhecimentos prévios esperado em cada uma das disciplinas, visto que na falta de tais informações, alunos acabam por solicitar disciplinas aos quais não estão preparados ou que são dos seus interesses, gerando como consequência o trancamento, ou até mesmo o abandono delas. A partir da elaboração de uma ontologia, conteúdos educacionais das disciplinas da matriz curricular dos cursos de engenharia podem ser representados e organizados, fornecendo a descrição dos conceitos e as relações existentes em um dado domínio para o compartilhamento e entendimento comum. A ontologia vem para ajudar a elucidar as escolhas dos alunos.

5.2 Objetivos Específicos

- Tornar o processo de aprendizagem autônomo e dinâmico.
- Elaborar uma ontologia das disciplinas presentes no campus.

6 Metodologia

A pesquisa bibliográfica será a ferramenta primária para a execução do projeto, abrangendo a compreensão do problema, elaboração da fundamentação teórica, e proposta da solução. O estudo do domínio e problema deve se basear no estudo das plataformas e soluções existentes, e na consulta às equipes e profissionais envolvido no contexto.

6.1 Organização Interna

O grupo foi dividido em diferentes frentes de trabalho, abrangendo o desenvolvimento da fundamentação teórica, estudo do contexto, e viabilidade da solução. A comunicação interna foi guiada por reuniões presenciais, com suporte de um mensageiro instantâneo para alinhamentos pontuais e menos formais.

As ferramentas utilizadas para o desenvolvimento do trabalho foram o Google Drive para gerenciamento de arquivos e criação de documentos, o *LaTeX* para a finalização dos documentos, e o Protegé para a modelagem da ontologia

6.2 Descrição da Solução

A proposta inicial dessa pesquisa é que o produto possa ser desenvolvido no **portal da FGA** (Faculdade do Gama) com integração com o **Matrícula Web**, onde estão dispostas as Grades Curriculares dos cursos da Universidade de Brasília.

A FGA é um campus da UnB(Universidade de Brasília) dedicado a desenvolver e formar conhecimentos e conhecedores das seguintes áreas: Engenharia de Software, Engenharia Automotiva, Engenharia de Energia, Engenharia Eletrônica e Engenharia Aeroespacial.

Os cursos oferecem cerca 560 (quinhentos e sessenta) vagas anualmente para interessados. A demanda por tecnologia inovação entre outras melhorias é grande, portanto a implantação de um portal de comunicação acadêmico onde pode-se encontrar desde informações aleatórias sobre a instituição até conhecimentos sobre TCC's apresentados pelos alunos, seus orientadores e as fontes de pesquisa corrente no campus.

Atualmente o portal encontra-se desenvolvido na plataforma *Noosfero*, que possui como base a linguagem *Ruby*. Esse software é de domínio público e é mantido de forma colaborativo em seu repositório *GitHub*.

Portanto esse ambiente selecionado apresenta demanda elevada no aspecto de melhoria de devolução de informações de forma dinâmica e assertiva aos interessados da comunidade acadêmica, sobretudo um campus de tecnologia sempre amplia o desenvolvimento de novas soluções. Assim acreditamos que um bom ponto de início é o **portal da FGA**, que é mantido por alunos da Engenharia de Software, e pelo professor Coordenador do curso de Engenharia de Software, Paulo Meireles aumentando ainda mais o potencial de evolução da pesquisa.

O proposto pela equipe é uma integração do **Matrícula Web** com o **portal da FGA**, onde os dados seriam obtidos utilizando o **Matrícula Web**, porém o sistema informatizado estaria presente apenas no **portal da FGA**. Ao acessar este sistema, os usuários poderiam obter informações que estão semanticamente representadas, enriquecendo a busca e garantindo maior eficiência de uso.

6.3 Desenvolvimento da Ontologia

A metodologia Desenvolvimento de Ontologias 101, apresentada por Natalya F. Noy e Deborah L. McGuinness [2], define um ciclo de vida para o desenvolvimento de uma nova ontologia. As atividades propostas cobrem aspectos importantes para a construção de uma nova ontologia, envolvendo os aspectos que, de forma geral, serão importantes no processo de definição e modelagem, a partir de um contexto.

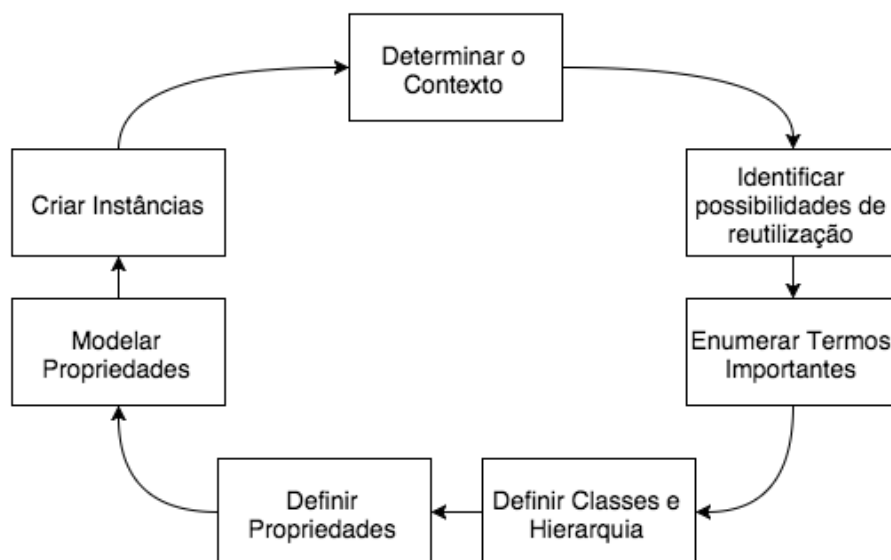


Figura 1. Processo de desenvolvimento

Esta ontologia não tenta definir um processo completo, mas sim propor um conjunto de atividades genéricas que, iterativamente, auxiliarão na representação do contexto trabalho em uma ontologia.

- **Determinar o contexto:** determinar e delimitar o domínio de trabalho, que será representado pela ontologia;
- **Identificar possibilidades de reutilização:** pesquisar por ontologias prontas, fazendo o possível para reutilizar e adaptar trabalhos existentes;
- **Enumerar termos importantes:** listar os termos mais importantes do domínio, os quais devem fazer parte da representação;
- **Definir classes e hierarquia:** transformar a lista de termos importantes em classes, e modelar os relacionamentos entre elas;
- **Definir propriedades:** listar os atributos que devem ser representados para cada uma das classes criadas;
- **Modelar propriedades:** definir como cada um dos atributos listados devem ser representados, e de que forma podem ser reaproveitados;
- **Criar Instâncias:** criar instâncias do mundo real para as classes de representação.

A utilização de tal metodologia neste trabalho deverá, além de ajudar na construção da ontologia, auxiliar na definição e construção do problema. As atividades propostas também serão importantes para a modelagem da solução, já que haverá um intenso estudo do domínio e dos conceitos fundamentais para a sua representação.

7 Descrição dos Usuários a Serem Beneficiados

Como o objetivo do projeto está no contexto de Universidades, mais especificamente, no contexto da Universidade de Brasília - UnB Gama, os usuários que terão contato com a solução são pessoas que frequentam a Universidade. Os Usuários e sua descrição podem ser observados na tabela a seguir:

Usuário	Descrição
Professor	Usuário que possui como interesse conhecer as relações entre ementas de disciplinas para poder garantir maior conhecimento sobre o contexto do conteúdo dado em sala. Conhecer essas relações garante que o professor poderá colaborar com alunos indecisos ou até utilizar isso para conhecimento próprio do conteúdo.
Aluno	Usuário que possui como interesse conhecer as Ementas, os Pré-Requisitos e as relações entre ementas das disciplinas. A utilização do sistema apoiará a escolha de disciplinas do Aluno, de acordo com seus interesses e importância.
Outros	Usuário que possui como interesse conhecer os Cursos, Ementas e e Relações entre disciplinas de um campus formado unicamente por Engenharias.

Tabela 2. Descrição dos Usuários

8 Descrição do Ambiente Web

O sistema do **portal da FGA** é um sistema na plataforma *Noosfero* desenvolvido e mantido por alunos e professores do curso de Engenharia de Software da Universidade de Brasília, dessa forma, a equipe do projeto possui grande suporte com relação ao **portal da FGA**, que é onde estará presente o sistema semanticamente representado. O Matrícula Web será a fonte dos dados utilizados neste sistema.

Sobre o **Matrícula Web**, a questão é um pouco mais complicada. O **Matrícula Web** é um sistema da Universidade de Brasília, desenvolvido e mantido pela Universidade, porém em outra instância, não fazendo parte do Campus Gama da UnB. Dessa forma, o Matrícula Web será utilizado apenas para obtenção dos dados, seja por meio de integração ou por meio de *Web Scraping*.

9 Cronograma do Projeto

O cronograma do projeto foi desenvolvido com apoio da ferramenta *Gantter*, e o mesmo pode ser acessado clicando [aqui](#).

<input type="checkbox"/> Determinar Domínio e Escopo	6d	09/09/2015
<input type="checkbox"/> Definir Tema do Trabalho	32h	09/09/2015
Definir Contexto	4d	09/09/2015
Definir Objetivos	4d	09/09/2015
Definir Frentes de Trabalho	2h	14/09/2015
Definir Cronograma	8h	16/09/2015
<input type="checkbox"/> Considerar Reutilização de Ontologias Existentes	6d	14/09/2015
Pesquisar Ontologias Relacionadas ao Tema	6d	14/09/2015
Executar Pesquisa Bibliográfica	6d	14/09/2015
Construir a Estrutura do Artigo	4h	16/09/2015
<input type="checkbox"/> Enumerar Termos Importantes na Ontologia	3d	21/09/2015
Listar Termos Importantes para o Contexto	3d	21/09/2015
Descrever o Problema	3d	21/09/2015
<input type="checkbox"/> Definir as classes e a hierarquia	5d	21/09/2015
Definir Entidades	3d	23/09/2015
Descrever situação atual do problema	3d	23/09/2015
Investigação do Domínio	5d	21/09/2015

Figura 2. Cronograma do Projeto

Referências Bibliográficas

- [1] BERNERS-LEE, T. et al. The semantic web. Scientific american, New York, NY, USA:, v. 284, n. 5, p. 28–37, 2001.
- [2] NOY, N. F.; MCGUINNESS, D. L. et al. Ontology development 101: A guide to creating your first ontology. [S.l.]: Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, 2001.