

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	3
	5 .1 Objetivos Gerais	3
	5 .2 Objetivos Específicos	3
6	Metodologia	4
	6 .1 Organização Interna	4
	6 .2 Descrição da Solução	4
	6 .3 Desenvolvimento da Ontologia	4
7	Descrição da Proposta	5
	7 .1 Contexto	5
8	Sentença de Posição do Produto	6
	8 .1 Usuários	6
	8 .2 Premissas	6
	8 .3 Restrições	6
9	Descrição das Ontologias Existentes	7
10	Descrição do Ambiente Web	7
11	Descrição da Solução	7
	11 .1 Ferramentas Utilizadas	7
12	Cronograma do Projeto	9

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. Além da especificação de um sistema que apoie a utilização da Web Semântica em *Ruby on Rails*.

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

No início, as páginas da web eram desenvolvidas por programadores de software, e todas essas páginas criadas eram informações direcionadas para 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 a cargo dos seres humanos.

O fato dos computadores não conseguirem interpretar essas informações está no fato de que as páginas da web não tem informação sobre si mesma, 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 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. Os alunos das Universidades Federais brasileiras, mais especificamente, os alunos da Universidade de Brasília possuem grande dificuldade em selecionar a grade de disciplinas a serem cursadas no semestre seguinte. Esta dificuldade só é possível graças a flexibilidade da matriz curricular dos cursos da UnB. A UnB exige que 70% do curso seja composto de disciplinas obrigatórias, formando a estrutura central do curso. Os 30% restantes podem ser distribuídos em disciplinas optativas a escolha do Aluno. Dessa forma, os próprios alunos desenham o seu curso a partir de seu interesse em determinadas áreas. Assim são formadas as ênfases durante o curso, por exemplo.

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.

Como o objetivo de apresentar o Problema a ser atacado de forma clara, foi utilizada a técnica do Diagrama de Ishikawa, que pode ser observada na imagem 1.

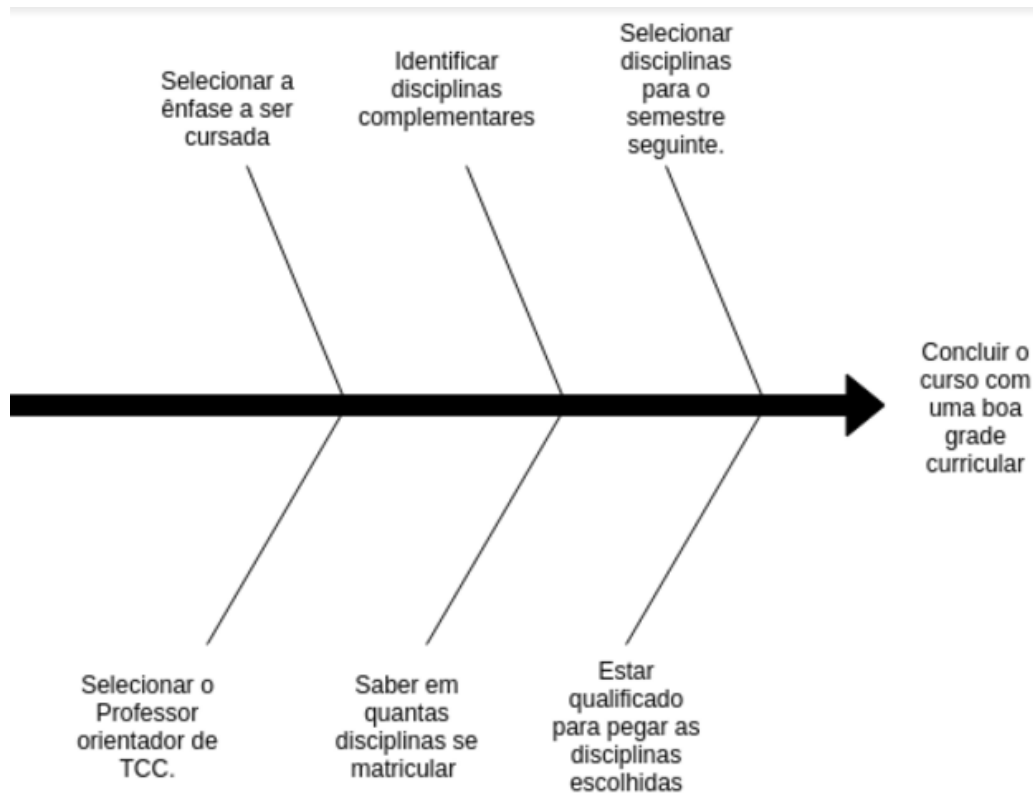


Figura 1. Diagrama de Fishbone

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 acabar por solicitar disciplinas aos quais não está preparado 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 organizadas, 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.

O projeto de construção do sistema de representação semântica no Portal FGA tem como objetivo verificar a viabilidade da utilização do Matrícula Web como fonte de dados sobre as disciplinas e cursos da UnB - Gama. Possui como objetivo, ainda, desenvolver uma arquitetura que apoie o sistema de representação semântica, utilizando Ruby on Rails.

A busca de ontologias existentes e referentes ao contexto estudado também é um dos grandes objetivos do projeto, já que primeiramente deve-se procurar ferramentas que auxiliem o projeto, caso não exista nenhuma, a única forma é desenvolver a ferramenta.

5.2 Objetivos Específicos

- Desenvolver a especificação de requisitos do sistema que apoiará a ontologia.
- Elaborar uma ontologia das disciplinas presentes no campus.
- Viabilizar a integração dos sistemas Portal FGA e Matrícula Web

- Buscar ontologias que possam embasar a construção da Ontologia proposta.

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.

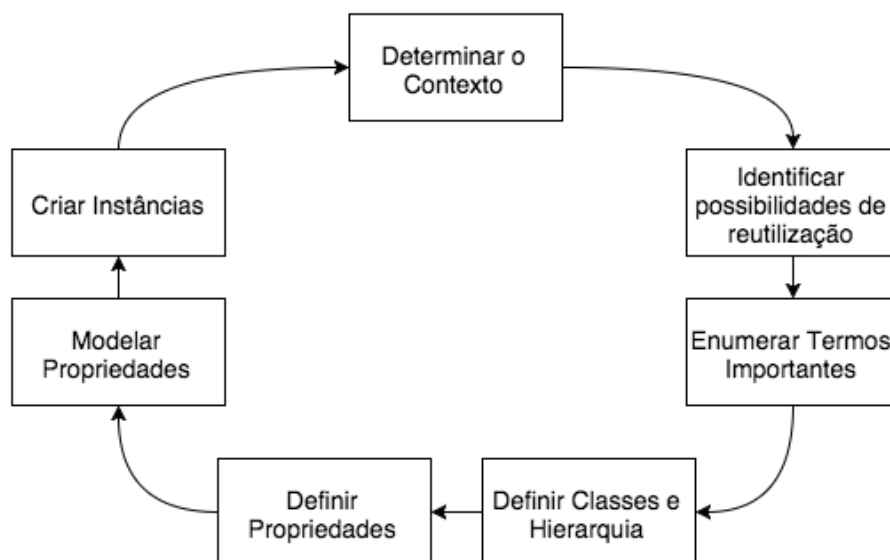


Figura 2. Processo de desenvolvimento

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 da Proposta

A solução proposta permitiria a representação semântica das informações contidas no Matricula Web, integrando às informações contidas no Portal da FGA. Além de garantir a conformidade com linked data e prosseguir com as iniciativas de web semântica dentro da Universidade de Brasília, seria possível relacionar cursos, departamentos, e Trabalhos de Conclusão de Curso de acordo com a área de estudo. Também seria possível possibilitar a criação de um sistema de aconselhamento de disciplinas, de acordo com o perfil do estudante.

7.1 Contexto

O sistema proposto funcionaria como um intermediário entre o **portal da FGA** e o <http://matriculaweb.unb.br> **Matricula Web**. O Portal reúne informações a respeito da Faculdade do Gama, incluindo informações gerais a respeito dos cinco cursos (Engenharia de Software, Engenharia Automotiva, Engenharia de Energia, Engenharia Eletrônica e Engenharia Aeroespacial), como por exemplo corpo docente e descrição. Por outro lado, o Matricula Web é a fonte primária de informações a respeito das ofertas, fluxos, e grades horárias.

O Portal é desenvolvido com base na plataforma Noosfero, que por sua vez foi desenvolvido em linguagem Ruby. O Noosfero é um software livre, e mais informações a respeito do Portal podem ser encontradas na instância do GitLab no e no **Portal do Software Público**. No entanto, é mais difícil encontrar maiores informações a respeito do Matrícula Web, que é uma interface para um sistema de gestão de matrículas mais antigo, e cuja base de dados é de acesso restrito dentro da Universidade de Brasília.

Existe uma demanda elevada no aspecto de melhoria de representação das informações, sobretudo um campus de tecnologia que amplia continuamente o desenvolvimento de novas soluções. No entanto, existe a problemática do acesso às informações do Matrícula Web, já que a base de dados é restrita. Por outro lado, a integração dos conceitos de web semântica na plataforma utilizada para o desenvolvimento do portal é promissora, já que se trata de um software, com uma comunidade ativa, e em constante evolução.

8 Sentença de Posição do Produto

Para	Todo o corpo discente da Universidade de Brasília.
Que	Deseja selecionar as disciplinas mais adequadas para o seu contexto.
Os	Alunos
Que	Selecionam disciplinas indesejadas ou até inúteis para o seu contexto de atuação.
Ao Contrário de	Selecionar apenas disciplinas que agreguem valor a sua formação.
Nosso Produto	Apresentará as melhores opções de matrícula para o Aluno, de acordo com seus interesses e contexto atual.

Tabela 2. Sentença de Posição do Produto

8.1 Usuários

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 3. Descrição dos Usuários

8.2 Premissas

- *Informações referentes às ofertas e fluxos estarão disponíveis para consulta:*
A fonte de pesquisa utilizada será o Matrícula Web a partir da utilização de *Web Scraping*, desse modo, a fonte de dados estará disponível sempre que o sistema do Matrícula Web estiver online.
- *Será possível adicionar funcionalidades ao Portal FGA:*
Como o Portal FGA é mantido pelos próprios alunos da Universidade de Brasília - UnB Gama, não existe muita burocracia para alteração ou inclusão de novas funcionalidades.
- *Será possível alterar a estrutura das informações da Universidade representadas no Portal FGA.*

8.3 Restrições

- Acesso limitado as informações;
- Escopo muito grande, sendo necessário primeiramente restringir à FGA;

- As informações a serem utilizadas pelo sistema são atualizadas com frequência, e afetam muito a funcionalidade.

9 Descrição das Ontologias Existentes

Existem iniciativas para a utilização de web semântica na representação de informações a respeito da estrutura organizacional de universidades, como por exemplo o **Linked Universities**. Esta organização lista as universidades no mundo que representam suas informações em formato linked data, e também todas os vocabulários hoje utilizadas para este tipo de representação, incluindo ontologias e fontes de terminologia.

Dentre as ontologias utilizadas, pode-se destacar a desenvolvidas por **Dave Reynolds**, que foca em estruturas organizacionais genéricas. Também, existe a estrutura proposta por **Rob Styles e Nadeem Shabir**, que ainda traz o ponto de vista organizacional, no entanto do ponto de vista de instituições acadêmicas, se aproximando mais do contexto proposto neste presente trabalho.

Adicionalmente, também vale indicar o trabalho desenvolvido por **Jeff Heflin**, que consiste em um rascunho de uma ontologia semelhante às anteriores e que ainda inclui informações a respeito de publicações.

10 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*.

11 Descrição da Solução

A solução proposta permitirá a representação semântica das informações contidas no Matrícula Web, integrando às informações contidas no Portal da FGA. Além de garantir a conformidade com linked data e prosseguir com as iniciativas de web semântica dentro da Universidade de Brasília, seria possível relacionar cursos, departamentos, e Trabalhos de Conclusão de Curso de acordo com a área de estudo. Também será possível possibilitar a criação de um sistema de aconselhamento de disciplinas, de acordo com o perfil do estudante.

A obtenção dos dados ocorrerá a partir da utilização de Web Scraping no Matrícula Web. A técnica de Web Scraping é o termo utilizado para descrever a atividade de extrair dados de sites, formatando-os da forma desejada e guardando no Banco de Dados. Esta técnica será utilizada no Matrícula Web por motivos de privacidade do sistema, nos deixando sem acesso aos dados de outra forma.

Já no contexto do Portal FGA, os alunos da Universidade de Brasília possuem total acesso ao sistema, já que o mesmo é desenvolvido e mantido por alunos e professores do curso de Engenharia de Software. O Portal FGA é desenvolvido em linguagem Ruby e framework Rails, que também seria a linguagem utilizada para o desenvolvimento da API de Web Scraping.

11.1 Ferramentas Utilizadas

Para o desenvolvimento de toda a solução, muitas ferramentas serão utilizadas. As mesmas estão descritas a seguir.

Protégé

É a ferramenta líder de engenharia de ontologias, um editor livre e um sistema de aquisição de conhecimento. Ele fornece uma interface gráfica de usuário para definir ontologias, também inclui classificadores dedutivos para validar se os modelos são consistentes e inferir novas informações com base na análise de uma ontologia. Como Eclipse, Protégé é uma estrutura para que vários outros projetos sugiram plugins.

Este aplicativo é escrito em Java e utiliza pesadamente balanço para criar a interface do usuário. Protégé tem recentemente mais de 200.000 usuários registrados e está sendo desenvolvido na Universidade de Stanford, em colaboração com a Universidade de Manchester e está disponível sob a Licença Pública Mozilla 1.1. (Dragan Gašević; Dragan Djurić; Vladan Devedžić (2009). Model Driven Engineering and Ontology Development (2nd ed.). Springer. p. 194. ISBN 978-3-642-00282-3.)

Ruby on Rails

Ruby on Rails é um *framework* livre da linguagem de programação *Ruby*, que promete aumentar velocidade e facilidade no desenvolvimento de sites orientados a banco de dados, uma vez que é possível criar aplicações com base em estruturas pré-definidas e com gerenciamento de memória automático. As aplicações criadas utilizando o framework Rails são desenvolvidas com base no padrão de arquitetura MVC(*Model-View-Controller*).

Personalização das Recomendações

De forma a oferecer recomendações mais precisas, espera-se poder treinar o sistema construindo perfis de preferência para cada usuário. Além de inferir relações de dependência a partir das disciplinas já cursadas, também serão usados questionários para definir preferências pessoais, que serão usadas como parâmetros adicionais para a recomendação.

O questionário abaixo é um exemplo de uma série de questões que podem ser utilizadas para a criação de tal perfil.

1. Selecione a disciplina que demonstra a sua área interesse dentro de Engenharia de Software:

- (a) Engenharia de Requisitos
- (b) Desenho de Software
- (c) Fundamentos de Sistemas Operacionais
- (d) Interação Humano-Computador

2. Qual sua maior dificuldade dentro de Engenharia de Software?

3. Quantos créditos você deseja pegar no semestre?

- (a) 0 á 16;
- (b) 16 á 22;
- (c) 22 á 26;
- (d) 26 a 32.

4. Qual sua disponibilidade de horário?

- (a) Matutino;
- (b) Vespertino;
- (c) Ambos.

5. Qual sua disponibilidade de tempo para estudo?

- (a) Pouco;
- (b) Razoável;
- (c) Muito.

6. Você gosta de trabalhar em equipe?

- (a) Sim;
- (b) Não.

7. Você é proavito?

- (a) Sim;
- (b) Não.

12 Cronograma do Projeto

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

☐ Determinar Domínio e Escopo	6d	09/09/2015
☐ 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
☐ 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
☐ 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
☐ 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 3. 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.

Apêndice

1. Entrevista Prof. Dr. Sergio Freitas

A questão do matricula web propriamente dito que cuida é o CPD, então o CPD mantém toda o banco de dados, toda linguagem, todo processo mesmo, todo desenvolvimento é CPD, não tem nada com nenhuma outra pessoa ou órgão na instituição. Basicamente o matricula web é uma ponta da parte acadêmica, e dentro da parte acadêmica há toda uma modelagem de banco de dados para modelar currículos, para modelar diversos projetos pedagógicos e etc. O que nós vemos no matricula web na verdade é a oferta referente a um dado período de tempo, com relação a um serie de disciplinas, casados, inicialmente, com um banco de professores, porque toda disciplina que está em determinada grade curricular ela vai ser alocada temporalmente em determinado momento para um professor e espacialmente para saber em qual sala está sendo dada a matéria.

Depois, no segundo momento, quando há a matricula efetivamente, aí aloca-se dentro desse slot da turma, uma turma é uma disciplina da grade, com um professor com um espaço físico e um determinado horário. Dentro da turma há o matricula web, que nada mais é do que, dado já essa oferta, as pessoas vão entrar ali dentro, so que cada pessoa para entrar no matricula web, que seriam os alunos, precisam também atender pré-requisitos, e os pré-requisitos estão relacionados as disciplinas que o aluno já fez, temporalmente em algum momento pra trás, dado o curso que ele está fazendo eu sei o que ele tem que fazer, então ele vai olhar e verificar quais são as disciplinas que ele já fez, se ele fez com aproveitamento, se não fez, quantas vezes ele fez, tudo isso está gravado no histórico daquele aluno, ai de tudo isso é feito um cálculo, pra poder saber se ele pode fazer ou não uma disciplina em que ele queira se pré-matricula. Depois quando ele atende esse pré-requisito, chega um segundo pré-requisito que é classificação, aí são outras coisas, coisas distintas.

Na minha opinião, o que vocês estão querendo fazer não está feito diretamente na matrícula web, está feito sim ao banco de dados que está por trás e que tem haver justamente com a modelagem do curso em si, qualquer curso que você está fazendo, porque dali que você quer na verdade, a única ponta que deixa margem a isso que vocês estão pensando, no meu entendimento, é a questão do TCC, porque o TCC acaba sendo uma instancia dinâmica e temporal que surge e que pode ser incorporada, não diretamente ao projeto pedagógico, mas na instancia de matricula. Apesar do no matricula web aparecer as informações, a formação daquilo não é ali e só vai servir para a fase final que é o aluno ir lá e falar, eu posso aqui eu posso ali. Então apesar da dificuldade da escolha que é real, o matricula web não trabalha direcionando para o perfil da pessoa, ele direciona para o perfil do curso. O que vocês estão pretendendo fazer é dizer o seguinte, eu gostaria de ser considerado, apesar de que existe algum tipo de consideração a nível do usuário, você não pode fazer isso porque você não tem pré-requisito, relações de dependência, pré-requisitos entre as disciplinas, mas isso é mais restrição, não é o contrário, mostrar o que seria bom você fazer, pelo menos ordena-las de tal forma que ela seja mais propícia para um dado perfil de aluno, ou seja, pra você, seria diferente pra ele, que seria diferente pra mim, mesmo que nós tivéssemos o mesmo leque de disciplinas disponíveis, identificação de perfil. Então vocês podem usar o material disponível do matricula web, no sentido de que, o que está lá é a oferta, tanto que seria só do semestre atual, você não consegue ver dos semestres anteriores, então aquilo não reflete exatamente o que é o curso de engenharia de software, porque só estão as ofertas desse período lá, começa por aí. Então, você pode pegar aquilo como rastro, para vocês usarem, para ficar mais fácil, mas tenham ciência de que aquilo não reflete o curso como um todo.

O matricula web é muito limitado em relação ao que vocês estão sugerindo, mas tem informações lá que são uteis para vocês fazerem estudos, não há como ter acesso ao banco de dados, para ter acesso é muito complicado, mas temos acesso a um rastro temporal disso que é um semestre, então pode-se usar isso para montar a ontologia. O trabalho de vocês quer estabelecer uma relação entre duas disciplinas existentes de tal que vocês possam aconselhar o aluno, dado um determinado, por isso que eu acho que dever ser por aluno e não só para o curso, para o curso, você tiver um trabalho de aconselhamento do curso, não dá para saber se é melhor ou pior, porque o projeto pedagógico foi construído baseado na cabeça de nos professora, que construímos o projeto pedagógico, baseado em outras experiências e ai se construiu

um ciclo de disciplinas, que tem um fluxo obrigatório, optativas. Então tratando ao nível de disciplina, introduzindo a variável aluno, que é fácil para vocês, porque vocês são alunos, então o cliente são vocês mesmo, fica mais tranquilo, então deveria inserir aí o ganho do aluno e aí fazer as perguntas que vocês gostariam de fazer para mim, que eu imagino que é, como fazer para fazer isso? Quais são as informações que preciso ter para ter uma abordagem? Provavelmente as perguntas que vocês deveriam me enviar ou vão me enviar são pautadas nesse contexto por isso: Eu tenho uma disciplina aqui, vamos pensar em duas disciplinas, e aí eu estou pensando em botar a visão de aconselhamento para esse aluno.

O problema seria eu ter um problema aqui embaixo e a pergunta que eu estou fazendo seria: eu fiz isso, eu tenho 3 disciplinas, quais delas é melhor eu fazer agora? Aí tem um problema de fácil compreensão e de importância crucial para vocês, agora o que eu tenho que fazer em termos de web semântica, e de ontologia de conceito. Bom primeiro ponto: no meu entendimento você tem que entender a modelagem da informação que o aluno vai fornecer para que você possa comparar com o que já existe. Porque um sistema de aconselhamento é o seguinte, não é só você falar assim, por que se você tirar o fator aluno eu vou ficar sempre aconselhando a mesma coisa. Isso não é um sistema de aconselhamento, é só uma página web estática já pronta que não aconselha nada. Então não muda nada de um semestre para o outro? Muda! Quais são as coisas que mudam, só os tópicos especiais? Não, muda ementa? Não a ementa não muda, mas muda professor! Por isso o aconselhamento seria sempre nesse sentido, não é bom porque não facilita para vocês... Apesar dos cursos da UnB possuírem no máximo 70% de créditos obrigatórios acabam não aproveitando muito bem isso, porque os alunos se sentem desconfortáveis em como usar esse 30%, o que causa esses problemas que vocês estão falando. Então, voltando, conhecendo as duas disciplinas, as ementas, as disciplinas que se ligam, qual seria a continuação natural? Isso já existe no Matricula Web para 1 semestre, e existe com força no projeto pedagógico que é a relação de dependência. Essa relação de dependência que parece ser obrigatória tem uma relação de conteúdo objetivo e se você aprendeu de forma objetiva o conteúdo da disciplina X1 você aprendera o conteúdo da disciplina X2, essa é a visão pedagógica.

E no caso do aluno que quer pegar disciplinas optativa de uma área específica? Conhecendo as disciplinas como estão hoje é muito difícil, de saber isso porque, ou elas estão elencadas formalmente *alguém já definiu quais optativas seguir para concluir sua ênfase*. Isso não é um aconselhamento com base em ontologia, é só um banco de dados com um filtro. Porém eu acho que esse não é o foco, se eu for para o grande plano, se você introduzir o perfil dos usuários, você pode pensar nos conteúdos que ele já fez, os conteúdos que ele se saiu melhor, por que você pode pensar o seguinte, comigo você não gosta da disciplina, mas você pode estar fazendo com o do lado e tirar nota alta, é normal! Então o que indica isso, que você tem mais afinidade com um do que com outro. Esse conteúdo teoricamente se você tivesse que indicar uma das duas disciplinas, tendo que você executou as duas já, o aconselhamento mostra a possibilidade de fazer as duas, isso é ruim, já que o aconselhamento deveria ser para a disciplina que você se saiu melhor. Então isso é só uma regra que deveria ser seguida, já é um indicador.

O outro indicador era você entrar dentro dos conteúdos das disciplinas, que vocês conhecem por ementa, e tentar olhar o texto e estabelecer uma relação com os textos das disciplinas seguintes. Como isso é feito? Na mão, ou você usa mecanismos automatizados para ver se as frases dos textos são relacionadas, aí você vai construindo uma ontologia de relacionamentos, mesmo que as duas disciplinas não estejam ligadas, porque não existe nada que diga que esta disciplina depende da outra, você está olhando para os conteúdos das duas e tentando achar um grau de relacionamento entre elas. Por exemplo eu acho que essa frase aqui se relaciona com essa frase, então a princípio você vai identificando quais seriam as disciplinas que tem um grau maior de relação, note que isso independe do aluno. Aí se você botar um peso nas duas frases ou 3 que identificou, você conseguiu categorizar essas frases que você identificou dentro daquilo que você quiser, exemplo segurança, se as frases estão categorizadas num banco de dados ou qualquer coisa aí você conseguisse categorizar automaticamente essas frases, aí você pode até ponderar se você olhar dá para disciplina dentro da UnB e der um pontinho a mais de acordo com o desempenho do aluno, aí você vai fazendo a ponderação.

Aí depois feita esse cruzamento a respeito da vida do aluno, faz-se o aconselhamento por conta do perfil do aluno, e não por conta da disciplina. Esse é um algoritmo por do qual vocês conseguiriam identificar usando web semântica e ontologias, como você aconselharia um aluno a fazer ou escolher disciplinas. Mas eu só aconselharia a fazer o aconselhamento do semestre seguinte, o percurso completo de todas as semestres é mais complicado, um algoritmo mais complexo!