

**Roteiro para elaboração de Relatório Técnico Individual Final (Bolsistas)**

O Relatório deve conter informações qualitativas e quantitativas, com enfoque tecnológico, obedecendo ao seguinte formato:

Identificação	Instituição: <b>Universidade Federal do Amazonas</b> Coordenador: <b>Raimundo da Silva Barreto</b> Nome do bolsista: <b>Fabício D'Morison da Silva Marinho</b> nº do Processo Institucional: <b>83.0091/2003-3</b> nº do Processo Individual/Área Prioritária: <b>13.1355/2008-2</b> Modalidade/Tipo de bolsa: <b>GM</b>
---------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**1. Conteúdo**

**Resumo do projeto:**

Descobrir automaticamente relações semânticas entre ontologias é uma tarefa de grande importância para Integração de Informação. Entende-se por **mapeamento** ou **alinhamento semântico** o problema de estabelecer analogias na forma de **axiomas ponte** entre conceitos de ontologias distintas, o que não é abordado pela maioria dos sistemas de alinhamento devido à dificuldade de tratar heterogeneidade semântica. Adotando a linguagem OWL, este trabalho objetiva explorar o problema de forma diferenciada, uma vez que recorre a Classificação Supervisionada: o método fundamenta-se na possibilidade de aplicar restrições lógicas comparando a instanciação dos conceitos das ontologias a mapear. Inicialmente, uma indução (**propagação bottom-up**) generaliza os valores de interseção e similaridade calculados entre conceitos durante a etapa de classificação. Sem perder a noção de direção na qual estes valores são computados, **regras de compatibilidade** entre conceitos definem mapeamentos de **equivalência, mais geral, menos geral, sobreposição e diferença**. Finalmente, estas regras são aplicadas seguindo uma **estratégia dedutiva top-down** capaz de computar mapeamentos mais confiáveis. O sistema desenvolvido chama-se *L-Match (Learning Match)*, é iterativo (pode reutilizar os mapeamentos computados) e utiliza diferentes algoritmos de classificação como sub-rotina, sendo notório o desempenho do classificador **Naive Bayes with Shrinkage (NB-Shrinkage)** que ajudou o *L-Match* a alcançar precisão e revocação acima de 80%. Uma abordagem auxiliar para seleção e desambiguação de sinônimos em vocabulários especializados foi desenvolvida com base em teoria de Grafos e da Informação, objetivando incrementar a precisão da classificação e conseqüentemente do mapeamento.

**Segmento/Atividade:**

Este projeto está inserido nas áreas de Inteligência Artificial e Recuperação de Informação para Integração de Informação. Especificamente é um projeto voltado à pesquisa de um sistema inteligente habilitado a descobrir equivalências e outros relacionamentos entre interpretações de mundo inseridas em bases de conhecimento distintas (ontologias) desenvolvidas por pessoas também distintas.

**Materiais e Métodos:**

As ontologias Ecolíngua (BRILHANTE, 2004) e Apes (ATHANASIADIS, at al., 2006) foram utilizadas para conduzir nossos experimentos. A Ecolíngua é uma ontologia do domínio ecológico, originalmente implementada em Prolog e posteriormente em OWL, instanciada com exemplos do PondSystem (GRANT, at al., 1997) (FORD, 1999) (HAEFNER, 1996). Por sua vez, a Apes é uma ontologia do domínio agrícola

desenvolvida pelo Projeto SEAMLESS-IF, relacionado a problemas de modelagem ambiental e agrícola. Outros pares de ontologias menores também foram utilizados incluindo ontologias sobre pontos turísticos da Rússia, ontologias sobre doenças e os catálogos de cursos/escolas da Universidade de Cornell (Austrália) e da Universidade de Washington (Estados Unidos). Além das ontologias, utilizamos o dicionário WordNet como fonte de sinônimos e a biblioteca *Rainbow* como fonte de algoritmos de classificação de texto. Por fim, o mapeador que nos permitiu rodar os experimentos foi implementado na linguagem JAVA e os experimentos consistiram em confrontar um conjunto de mapeamentos conhecidos por especialistas contra um conjunto de mapeamentos sugeridos automaticamente pelo mapeador desenvolvido.

Palavras-chave: integração de informação, ontologia, mapeamento semântico, aprendizado de máquina, Naive Bayes Shrinkage.

---

## 2. Principais Resultados   Resultados obtidos em função do projeto proposto

- Criação de bibliotecas de código para pre-processamento, experimentos de mapeamento e avaliação dos mapeamentos entre ontologias desenvolvidas na linguagem OWL.
- Resultados experimentais numéricos melhores que aqueles de trabalhos relacionados oriundos da literatura recente.
- Dissertação escrita sobre os trabalhos de pesquisa desenvolvidos durante o curso de mestrado.

### Conhecimentos adquiridos:

Adquiri conhecimentos sobre o que são ontologias e sua utilidade, como desenvolvê-las na ferramenta Protégé, estudei tópicos sobre *description logics* (DL), e aprendi bastante sobre técnicas de mapeamento, casamento e integração de conceitos de ontologias, de bases de dados e de outros tipos de esquemas. Também conheci outros algoritmos de classificação de texto como *naive bayes with shrinkage* e o algoritmo da máxima entropia.

## 3. Razão para interrupção da bolsa

O curso de mestrado foi completado, defendido e finalizado. O ex-bolsista está elaborando um artigo sobre os seus resultados para ser publicado em conferencia reconhecida.

Manaus, 07 de janeiro de 2010.

---

**Fabricio D'Morison da Silva Marinho**  
Bolsista

---

**Raimundo da Silva Barreto**  
Coordenador