

PROCESSAMENTO DE LINGUAGEM NATURAL EM TEXTOS DA FARMACOVIGILÂNCIA

ALUNO: ANDRÉ ALVES DE ARAÚJO

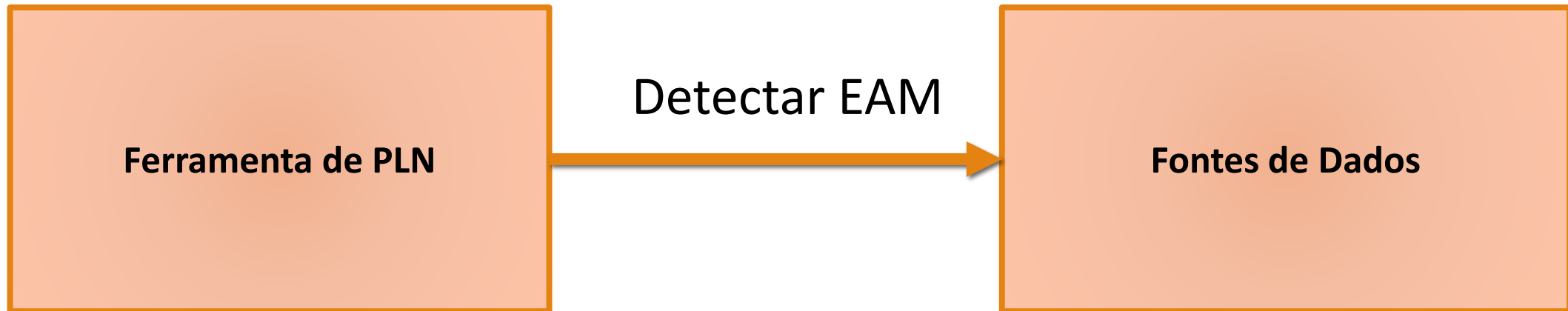
ORIENTADOR: GUSTAVO PAIVA GUEDES E SILVA, KELE BELLOZE*

Sumário

1. Introdução
2. Metodologia
3. Resultados
4. Discussão
5. Conclusão

Introdução

- Farmacovigilância
 - Eventos adversos relacionados a medicamentos(EAM)
 - Processamento de Linguagem Natural(PLN)
- Processo da Farmacovigilância



Introdução

- Metamap
 - UMLS
 - Modelos de Dados



- Problema
 - Mapeamento de conceitos apenas em inglês
 - Indiferente da linguagem do texto
 - Uso de modelos em Inglês

Introdução

- Objetivo
 - Permitir o mapeamento em português de textos em português
- Criação de um modelo de dados em português
- Testes no servidor e numa aplicação em Java
- Uso no contexto da farmacovigilância
 - Conteúdo (Post's) de mídias sociais

The logo for MetaMAP, featuring the word "MetaMAP" in a bold, red, sans-serif font. The letters have a slight 3D effect with black outlines. A registered trademark symbol (®) is located at the bottom right of the letter "P".

Metodologia

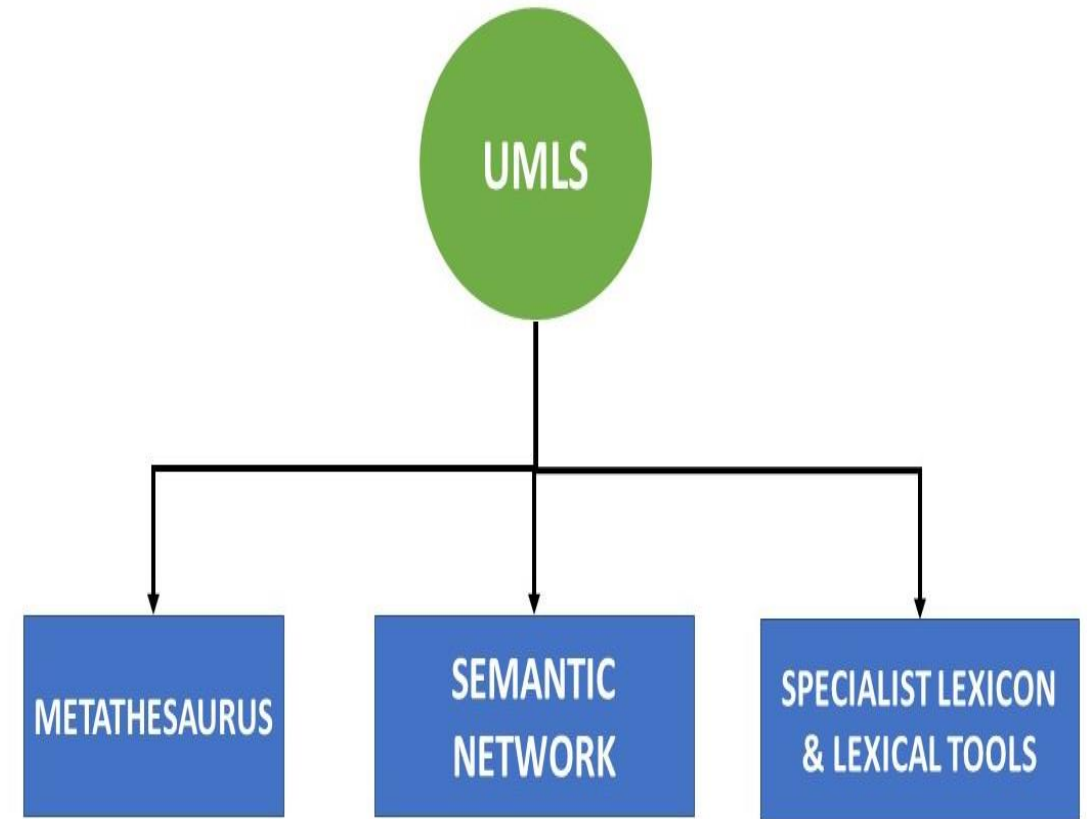
1. Metamap e UMLS
2. Processamento teórico do Metamap
3. Modelo de dados em português do Metamap
4. Uso do Modelo na aplicação em Java

The logo for MetamAP, featuring the word "MetamAP" in a stylized, bold, red font with a black outline. The letters are slightly slanted and have a 3D effect. A registered trademark symbol (®) is located at the bottom right of the letter "P".

MetamAP®

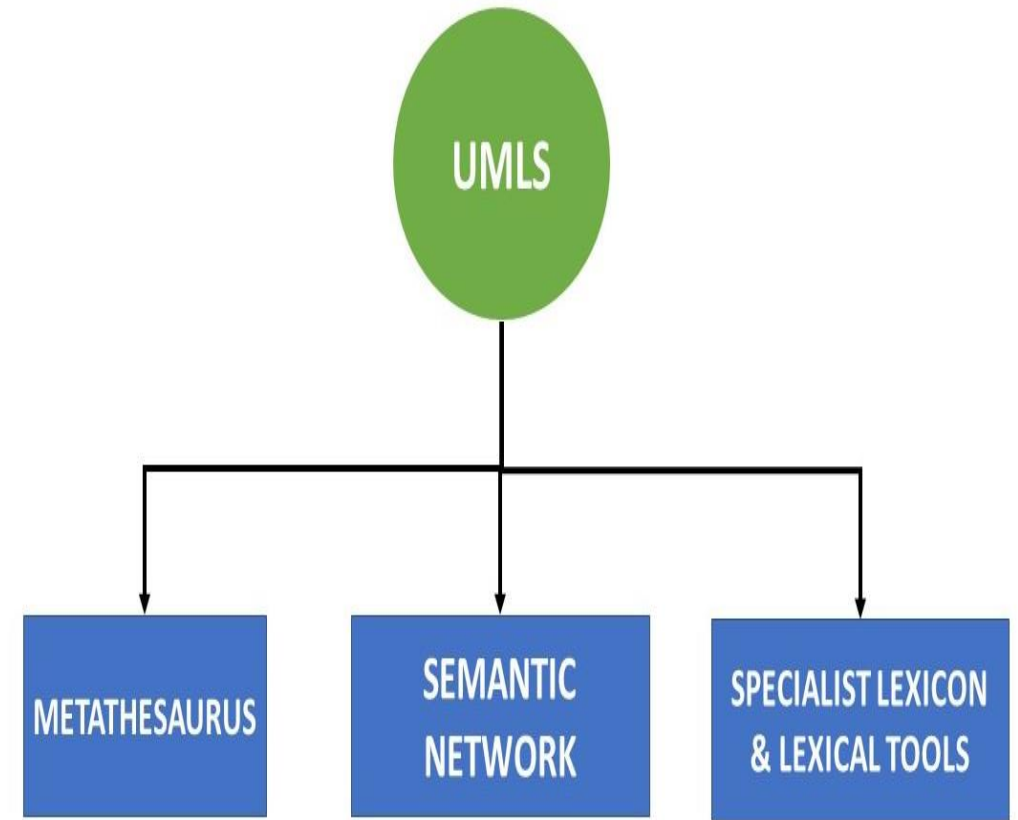
Metamap e UMLS

- UMLS
 - Metathesaurus
 - DOR → Dor, DOR, dor, ...
 - Semantic Network
 - Clinical Drug treats Disease or Syndrome
 - SPECIALIST Lexicon & Lexical Tools
 - DOR -> dores, dolorido



Metamap e UMLS

- Relacionamento com Metamap
 - Identifica os conceitos no Metathesaurus
 - Avalia os conceitos
 - Retorna os melhores conceitos médicos
 - String
 - Grupo semântico
 - Vocabulário



Processamento teórico do Metamap

1. Pré-Processamento

- Define tags sintáticas a partir do texto
- Analise sintática a partir dessas tags
- Expansão com tokens
- Limpeza de informações
- Correspondência com as palavras do texto

2. Geração de Variantes

- Gera todas as variações flexionadas e ortográficas das palavras de entrada
- Utiliza o léxico SPECIALIST

Processamento teórico do Metamap

3. Busca de Candidatos

- Recupera todas as strings Metathesaurus com suas informações principais (como ID, grupo semântico, nome do conceito) que começam com uma das variações criadas

4. Avaliação de Candidatos

- Calcula o Score de cada candidato encontrado.
- Paramêtros: Centrality, Variation, Coverage, Cohesiveness, Involvement

Processamento teórico do Metamap

5. Construção de Mapeamento

- Escolhe a melhor combinação de candidatos que formam o mapeamento mais completo possível (maior score)

Don't forget to get a diabetes screening

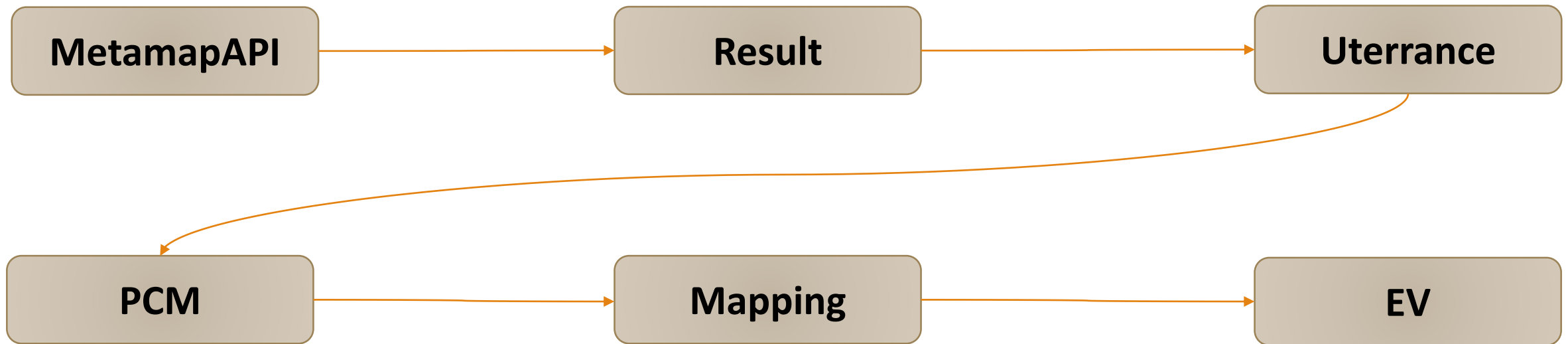
- ❑ forget
 - 1000 forgetting (forget) [Mental or Behavioral Dysfunction]
- ❑ to
 - 1000 Togo (TO) [Geographic Area]
 - 1000 Tryptophanase (TO) [Amino Acid, Peptide, or Protein.Enzyme]
 - 1000 To (To) [Qualitative Concept]
- ❑ a diabetes screening
 - 1000 Screening for diabetes (diabetes screening) [Diagnostic Procedure]

Modelo de dados em português do Metamap

- Estrutura de dados que permite ao servidor acessar os conceitos do UMLS de vocabulários específicos durante o processo de mapeamento
- Modelos em Inglês: USABase, NLM, Base
- Data File Builder
 - Permite criar um novo modelo de dados do Metamap
 - Vocabulários em Português do UMLS
 - ICPCPOR, LNC-PTBR, MDRPOR, MSHPOR, WHOPOR

Uso do Modelo na aplicação em Java

➤ Metamap Java API



Resultados

- Limitações do modelo de dados proposto:
 - Não aceita palavras com acentuações ou caracteres especiais fora da tabela ASCII
 - Alta chance de variações gramaticais não serem mapeadas corretamente. Exemplo:
 - dor → DOR
 - dolorido, dores -> sem conceito

Discussão

- Caracteres especiais fora da tabela ASCII
 - Regra do servidor independente do modelo de dados
- Variações gramaticais e o léxico inglês SPECIALIST
 - Utilizam regras da língua inglesa
 - Resultado em variações que são misturas de português e inglês que não podem ser mapeadas
- Metamap e a busca de EAM em posts de mídias sociais
 - Procurar grupos semânticos
 - Mapear gírias a conceitos médicos

Conclusão

- Novo modelo de dados em português do Metamap
- Documento Tutorial
 - Aspectos técnicos da criação e uso do modelo em português
 - Uso no servidor e aplicação Java
- Formas de resolver as limitações
 - Replace UTF8
 - Novo léxico médico em português
- Para auxiliar a farmacovigilância deve-se verificar o uso do Metamap com outros softwares.

PROCESSAMENTO DE LINGUAGEM NATURAL EM TEXTOS DA FARMACOVIGILÂNCIA

ALUNO: ANDRÉ ALVES DE ARAÚJO

ORIENTADOR: GUSTAVO PAIVA GUEDES E SILVA, KELE BELLOZE*

AGRADECIMENTOS: CEFET/RJ-PIBIC