

**SimpleNLP, uma biblioteca para
processamento de linguagem natural,
aplicada a análise de características
psicológicas.**

...

Motivação

- Falta de uma biblioteca para Processamento de Linguagem Natural simples e altamente customizável.
- Estudo das técnicas de processamento de linguagem natural
- Estudo do relacionamento entre linguagem e psicologia
- Experimentação de PLN para análise de aspectos psicológicos

Objetivos

- Implementação de técnicas relevantes de Processamento de Linguagem Natural em JAVA e Scala.
- Compilação da implementação em uma biblioteca chamada SimpleNLP
- Obtenção de base de dados relacionadas a aspectos da psicologia
- Experimentação dos métodos desenvolvidos na biblioteca SimpleNLP nas bases de dados coletadas

Linguagem

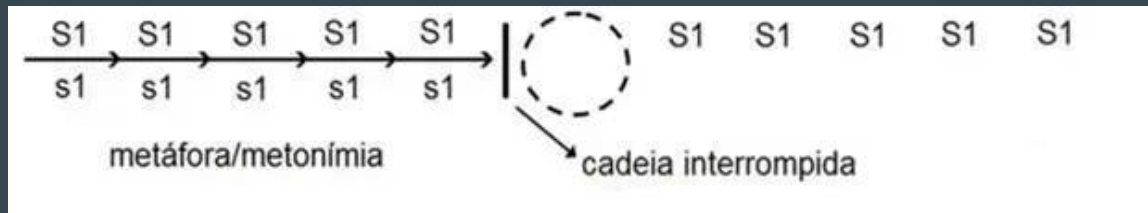
- A linguagem é tida como a qualidade inata ao ser humano de se comunicar através de uma língua
- A língua um meio concreto para comunicação.

Teorias filosóficas semânticas da linguagem

- Ideacionais
- Representacionalistas
- Pragmáticas
- Inferenciais

Linguagem e Psicologia

- Segundo Lacan a linguagem estrutura o subconsciente.
- Linguagem é um mapa para o real.
- Metáfora - Relação de semelhança entre significantes
 - "Essa menina é uma flor"
- Metonímia - Relação de continuidade entre significantes
 - “Lavou os cristais da avó antes de usá-los no jantar”

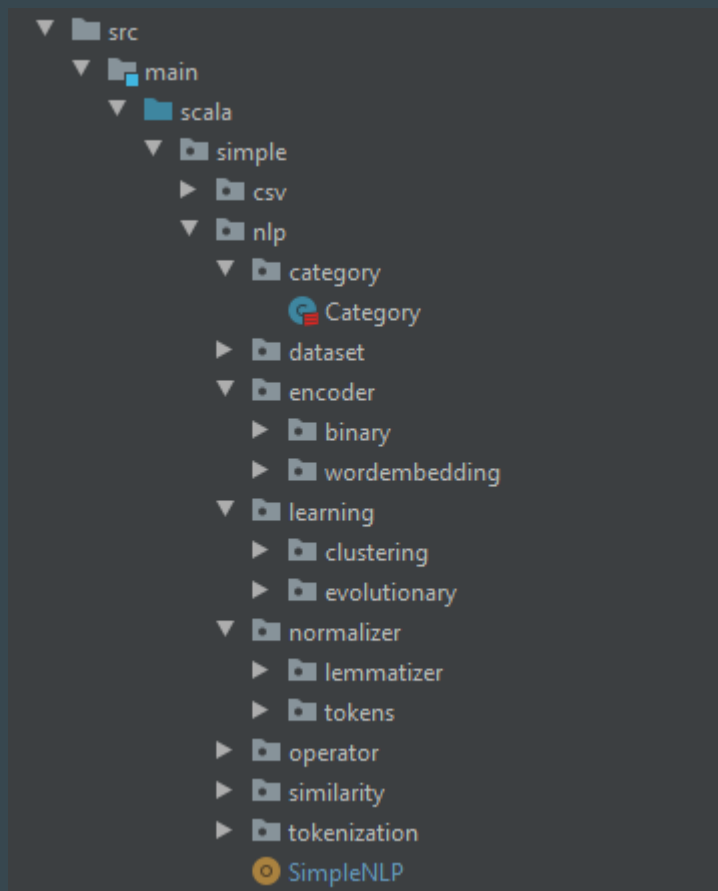


Processamento de Linguagem Natural

- Tokenização
- Similaridade de Cossenos
- Binary Encoding
- Word Embedding

Implementação

- JAVA e SCALA
- Apache Spark MLlib



Clustering

```
package nlp.learning.clustering

import simple.nlp.learning.clustering.LearningLloyds

object LearningLloydsTest {

  def main(args: Array[String]): Unit = {

    LearningLloyds.initModel( numOfClusters = 10, numOfDimension = 10)

    var inputs: Array[Array[Float]] = Array.empty
    for (num <- 1 to 10) {
      inputs = inputs :+ Array.fill[Float](10)(num.toFloat)
    }

    var test = LearningLloyds.optimize(inputs, iterations = 100)

  }
}
```

Algoritmo Genético

```
package nlp.learning.evolutionary

import simple.nlp.learning.evolutionary.LearningGenetic

object LearningGeneticTest {

  def main(args: Array[String]): Unit = {
    LearningGenetic.initModel( generationSize = 9, chromosomeSize = 10)

    //MUTATION IN 10%
    LearningGenetic.setMutationChance(10)

    //MANTAIN BEST 3 INDIVIDUALS IN GENERATION
    LearningGenetic.setElitismSize(3)

    //EXECUTE 2 CROSSOVER WITH THE ELITE
    LearningGenetic.setCrossoverSize(2)

    val input:Array[Float] = Array(1,5,6,5,7,4,8,1,4,3)
    var testSingleInput = LearningGenetic.optimize( objective = 1,input, iterations = 200);

    val mutipleInput:Array[Array[Float]] = Array.fill(10) (input)
    var testMutipleInputs = LearningGenetic.optimize( objective = 1,multipleInput, iterations = 200);

  }

}
```

Latent Semantic Indexing

```
object EncoderLSITest {  
  
  def main(args: Array[String]): Unit = {  
  
    var inputs: Array[Vector] = Array.empty  
  
    inputs = inputs :+ Vectors.dense( firstValue = 1, otherValues = 0,1,0,1,1,1,1,1,0,0)  
    inputs = inputs :+ Vectors.dense( firstValue = 1, otherValues = 1,0,1,0,0,1,1,0,2,1)  
    inputs = inputs :+ Vectors.dense( firstValue = 1, otherValues = 1,0,0,0,1,1,1,1,0,1)  
  
    var test = EncoderLSI.encode(inputs, numOfDimensions = 2)  
  
    var query: Array[Double] = Array(0,0,0,0,0,1,0,0,0,1,1)  
    var testQuery = EncoderLSI.encodeQuery(query)  
  
  }  
  
}
```

Experimentação

- Base de Tweets
- Aplicação do algoritmo LSI
- Querys “sad” e “happy”

Resultados

- Menos similar a happy: "- i think i need to find better anti-depressants i think this paxil/wellbutrin combo is losing its efficacy",
- Mais similar a expressão sad : "that's sad",
- "A retórica do depressivo, inserta no contexto discursivo, carece de respostas ao traduzir as histórias de vida, considerando como relevante sua identificação aos aspectos da doença e do doente". (Neto, Modesto Leite Rolim. "O Discurso e as Narrativas na Vivência da Depressão." Psicologia, Saúde & Doenças 6.2 (2005): 131-138.)

Conclusão

- A retórica do depressivo carece de termos relacionados a felicidade
- Talvez para trabalhos futuros seja interessante pensar em implementações de modelos de word embedding para análise de documentos menos relacionados a felicidade com o intuito do diagnóstico da depressão.