

---

## Errata

---

# Identificação automática de relações multidocumento

**Erick Galani Maziero**

**Orientador: Prof. Dr. Thiago Alexandre Salgueiro Pardo**

Este documento contém correções da versão  
revisada da Dissertação apresentada ao Instituto de  
Ciências Matemáticas e de Computação - ICMC-  
USP

**USP – São Carlos**

**Maio de 2012**

A seção 5.2.5.2. intitulada “Classificadores Multirótulo” contida nas páginas 61 a 63, tem seu texto, figura e tabelas substituídos pelo conteúdo a seguir.

### 5.2.5.2. Classificadores Multirótulo

Tendo em vista a sobreposição das classes, explorou-se uma metodologia de identificação multirótulo. Essa metodologia consiste em identificar os pares de sentenças que têm mais de uma relação CST (no máximo 2, como pode ser visto pelos rótulos gerados na Tabela 15) e gerar um novo rótulo contendo o nome das duas relações. No entanto, essa metodologia leva a um maior desbalanceamento das classes. A Figura 1 mostra a quantidade de cada classe (rótulos) obtida no processamento do *córpus* para obter os multirótulos.

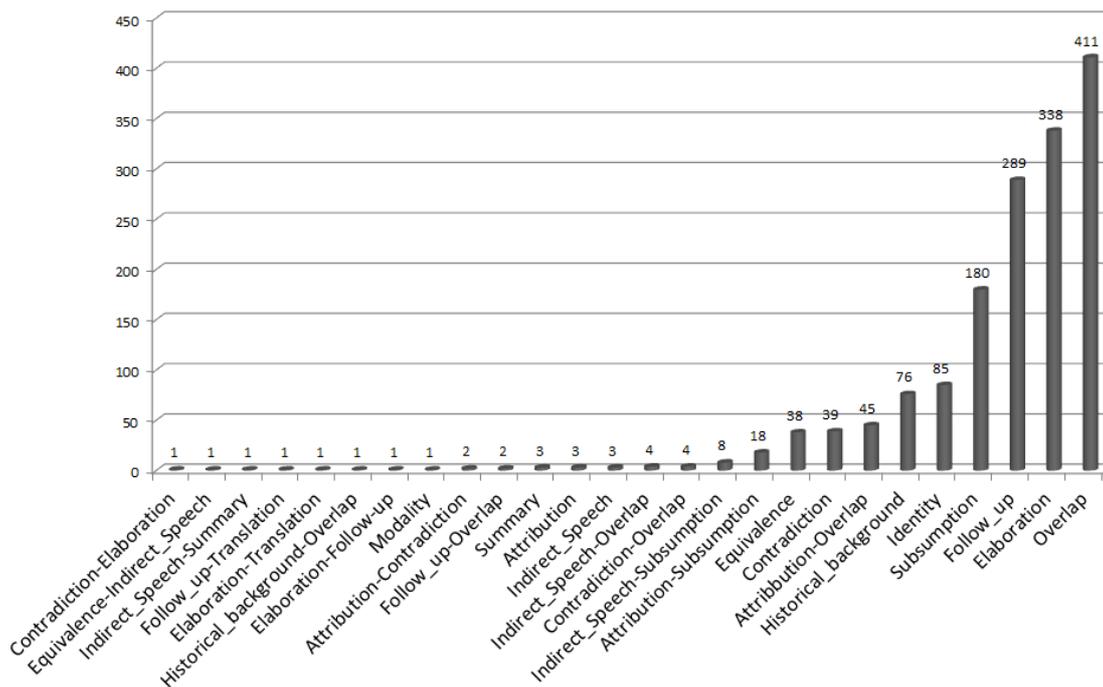


Figura 1 - Frequência das classes no classificador multirótulo

Os resultados da classificação multirótulo estão na Tabela 15. Vê-se que as novas classes criadas, contendo mais de uma relação CST, ficaram com valores nulos, exceto a classe “*attribution-overlap*” e “*subsumption-follow-up*”, que tiveram uma frequência suficiente

para algum aprendizado. Assim, a abordagem multirótulo, da forma como foi realizada, é inviável dadas as características do único corpus disponível para o aprendizado.

**Tabela 1 - Resultados do classificador multirótulo**

<b>Relação</b>	<b>Precisão</b>	<b>Cobertura</b>	<b>Medida-F</b>
<i>Attribution-Overlap</i>	0,103	0,156	0,124
<i>Contradiction_Elaboration</i>	0	0	0
<i>Attribution-Subsumption</i>	0,077	0,056	0,065
<i>Attribution-Contradiction</i>	0	0	0
<i>Indirect_Speech-Overlap</i>	0	0	0
<i>Equivalence-Indirect_Speech</i>	0	0	0
<i>Indirect_Speech-Subsumption</i>	0	0	0
<i>Indirect_Speech-Summary</i>	0	0	0
<i>Follow_up-Translation</i>	0	0	0
<i>Elaboration-Translation</i>	0	0	0
<i>Historical_background-Overlap</i>	0	0	0
<i>Contradiction-Overlap</i>	0	0	0
<i>Follow_up-Overlap</i>	0	0	0
<i>Elaboration-Follow_up</i>	0	0	0
<i>Subsumption</i>	0,380	0,406	0,392
<i>Elaboration</i>	0,390	0,388	0,389
<i>Follow_up</i>	0,305	0,325	0,315
<i>Overlap</i>	0,405	0,416	0,411
<i>Historical_background</i>	0,214	0,158	0,182
<i>Contradiction</i>	0,154	0,103	0,123
<i>Identity</i>	0,879	0,941	0,909
<i>Equivalence</i>	0,130	0,079	0,098
<i>Modality</i>	0	0	0
<i>Summary</i>	0	0	0
<i>Attribution</i>	0	0	0
<i>Indirect_Speech</i>	0	0	0
<b>Media</b>	<b>0,117</b>	<b>0,116</b>	<b>0,116</b>

Neste cenário, a técnica J48 ganha tanto do NaiveBayes quanto do SVM (Tabela 2) e é utilizada na criação do classificador multirótulo.

**Tabela 2 - Resultado do teste de significância estatística**

<b>NaiveBayes</b>	<b>SVM</b>	<b>J48</b>
0.3274	0.3495 (e)	0.3765 (g)