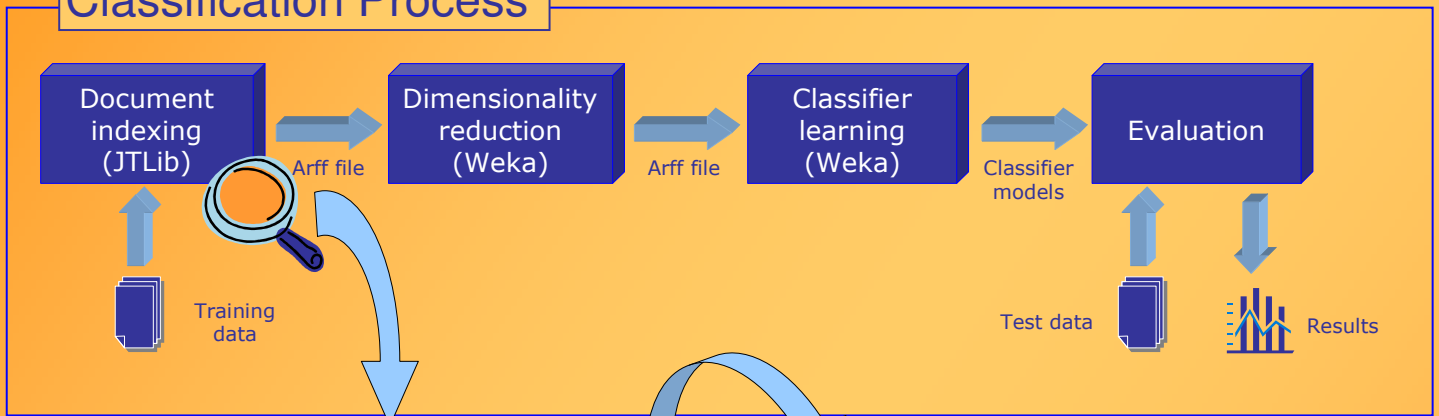


# Feature Engineering and Quick Prototyping of Gene Mention Classifiers\*

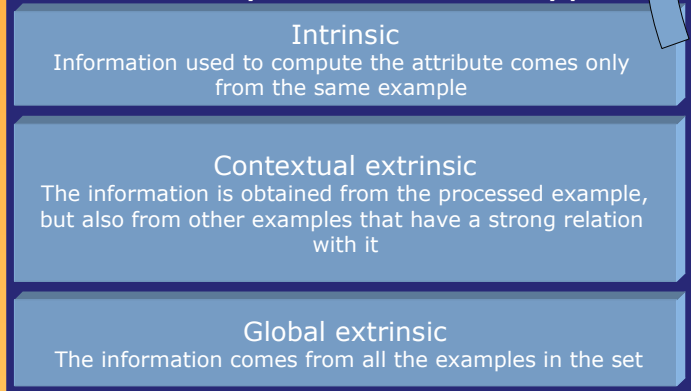
Manuel J. Maña López  
 Jacinto Mata Vázquez  
 Universidad de Huelva  
 {manuel.mana, mata}@diesia.uhu.es

Francisco Carrero García  
 José M<sup>a</sup> Gómez Hidalgo  
 Universidad Europea de Madrid  
 {francisco.carrero, jmgomez}@uem.es

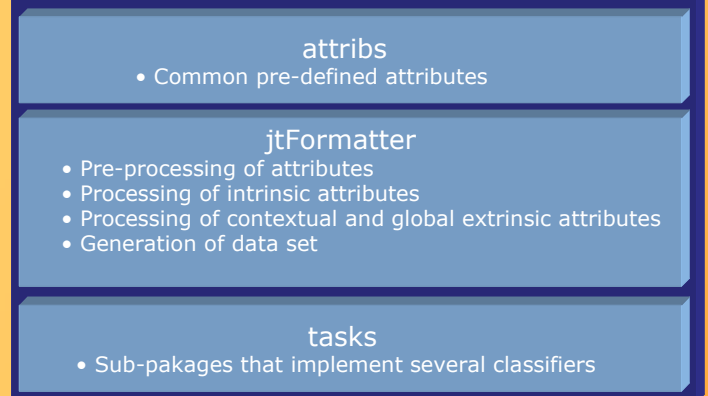
## Classification Process



## Taxonomy of Attribute Types



## JTLib



## Selected Attributes

Attribute type	Attribute Name
<b>Intrinsic</b>	hyphen
	punctuation
	initCaps
	lettersAndDigits
<b>Global extrinsic</b>	number
	frequentWords
	frequentWordsInEntity
	prev1Unigrams
	prev2Unigrams
<b>Contextual extrinsic</b>	startingWords
	endingWords
	endOfSentence
	punctuation ± <i>n</i>
	initCaps ± <i>n</i>
frequentWords ± <i>n</i>	
frequentWordsInEntity ± <i>n</i>	

## Ranking of Attributes (IG)

Attribute		Attribute	
1	frequentWords	8	frequentWordsInEntity + 1
2	frequentWordsInEntity	9	lettersAndDigits
3	frequentWords - 1	10	startingWords
4	frequentWords + 1	11	frequentWords - 2
5	endingWords	12	frequentWords + 2
6	frequentWordsInEntity - 1	13	frequentWordsInEntity - 2
7	prev1Unigrams	14	prev2Unigrams

## Results

	C4.5 unpruned	C4.5 pruned
<b>Precision</b>	50.09	53.37
<b>Recall</b>	46.12	42.46
<b>F-measure</b>	48.02	47.29