

# Text Categorization:

Using a linguistic approach  
for higher accuracy

# Keyword matching approach

"I like my new car because it's a hybrid"

—————  
Categorized as: "Customer  
satisfaction"

Rule followed: Categorize as "customer satisfaction" phrases containing: "like" and  
"car"

**Correct categorization**

# Keyword matching approach

"My neighbor's **car** is **like** mine and sometimes I get confused"

—  
Categorized as: "Customer  
satisfaction"

Rule followed: Categorize as "customer satisfaction" phrases containing: "like" and  
"car"

## Wrong categorization

If we categorize this phrase automatically, we can easily see that even if it contains the two words in the rule it's not related to the category we are analyzing

# POS approach

"One of the things I liked the most was using the App to book, I could even choose the floor I wanted "

---

Categorized as: "Booking service"

Rule followed: Categorize as "booking service" phrases containing: "like" when it's used as a verb and "book" when it's used as a verb

**Correct categorization**

# POS approach

I really **like** the latest **book** I've read, I finished it  
during the weekend "

---

Not Categorized as: "Booking service"

Rule followed: Categorize as "booking service" phrases containing: "like" when it's used as a verb and "book" when it's used as a noun

## Correct categorization

Even if like it's a verb, in this sentence book is used as a noun, so the rule it's not applied"

# How to use POS approach to create a coding plan

Create a category and establish all the POS  
you are interested in:

Category	POSeS followed
pricing	accessible;adj,price;sust
pricing	affordable;adj
pricing	cheap;adj
pricing	cheaper;adj
pricing	discount;adj
pricing	discount;verb
pricing	expensive;adj
pricing	extortionate;adj
pricing	inexpensive;adj
pricing	price;sust
pricing	pricey;adj
pricing	pricy;adj
pricing	value;sust,money;sust
pricing	worth;sust