CACTUS: A User-friendly Toolkit for Semantic Categorization and Clustering in the Open Domain

Open-domain
No Training Required
Easy-to-use GUI

Near-universal coverage
Internet as Knowledge Source
Or, command-line interface

CACTUS: A 1-Slide Introduction

Named Entities/Common Nouns
JetBlue
Microsoft
Angela Merkel

Categories
company
leader
plant

Category examples
Mattel
Taro Aso
hosta

"Companies such as Mattel and JetBlue"
327 hits

"Leaders such as Taro Aso and JetBlue"
18 hits

"Plants such as hosta and JetBlue"
2 hits

"___ such as ___ and jetBlue"

Categorizer: Pick the best category for each noun.
Clusterer: Cluster all nouns based on what categories they can belong to.