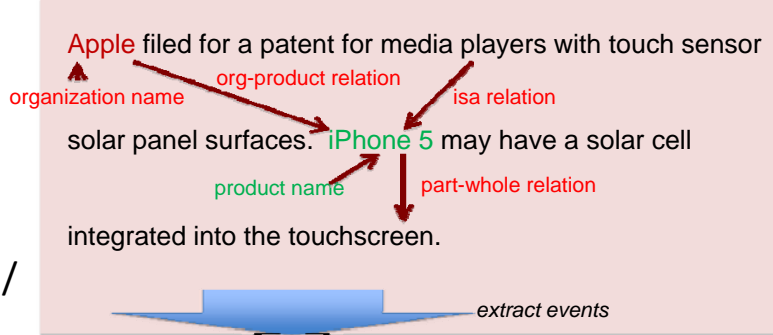


Proteus Project
New York University
<http://nlp.cs.nyu.edu/>

Providing Multilingual Information Extraction With Minimal Training Data



file-patent	
Agent	Product
Apple	media player ...

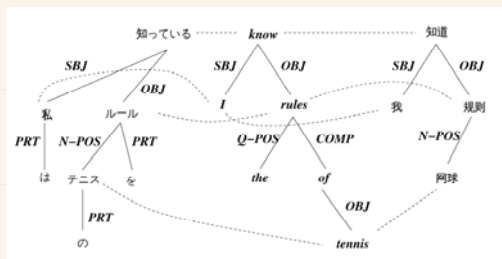
integrate		
Agent	Component	Product
Apple	solar cell	touchscreen

Capabilities and Advances in Information Extraction

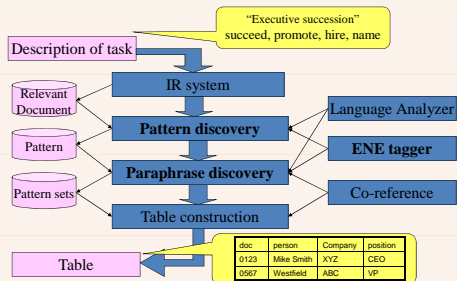
	Supervised	Semi-supervised	Unsupervised
Names	Extended named entity set (200 types) [Sekine]	Learn domain-specific classes from co-occurrence information [Lin; Min]	Identify names from diachronic term frequency info [Shinyama]
Relations and attributes	Kernel methods for building ACE relation models [Zhao]	Enhanced bootstrapping methods for ACE relations [Sun]	Identify primary relations between entity types by clustering [Hasegawa]
Events	MaxEnt models extended to use -cross-document information [Ji] -cross-event information [Liao]	Using document relevance for bootstrapping [Yangarber; Sudo] On-Demand Info. Extraction [Sekine]	"Pre-emptive IE" clusters events to create an inventory of event types in a corpus [Shinyama]

Extraction development in English, Chinese, and Japanese

GLARF:
Parallel multilingual
Predicate-argument
Representations for
English / Chinese / Japanese
[Meyers]



ODIE: On-Demand Information Extraction
Create a new IE system for a user-specified topic/event in a minute [Sekine]



Rapidly-retargetable info extraction can aid in observing trends, an essential component of technology tracking for FUSE.

Ralph Grishman (Project Director)
grishman@cs.nyu.edu
Adam Meyers
meyers@cs.nyu.edu