

Information Extraction.

① Obtain structured text/ from unstructured data.

Objectives: IE system.

- ① Find & understand limited relevant parts of texts
- ② Gather info from many different texts.
- ③ Produce a structured representation of the relevant info gathered.

Relations (Database)

Knowledge base / knowledge graph.

End goals.

- Make the data "useful" for users
- Organize the info in a semantically precise manner.

Input (unstructured / semi-structured data) $\xrightarrow[\text{by}]{\text{used}} \text{computer algorithms.}$ \rightarrow structured data.

In 1998 Larry Page and Sergey Brin founded Google Inc.

Founder Of (Larry Page, Google Inc.)
Founder Of (Sergey Brin, Google Inc.)
Founded In (Google Inc., 1998).

Applications?

Biomedical domain.

- ① Large no. of scientific publications from the biomed domain
- ② Discover easily innovations (ii) particular genes (ii) proteins & other entities.
- ③ Biomed entities ← ① many ambiguous names.
② lot of ~~synonyms~~ synonymous words.

Automatically identify the bio medical
entities and link. them in

Knowledge bases.

→ WikiData.

↓

subject	relation	object
p53	is-a	protein
p53	has function	apoptosis
apoptosis	involved-in	cell death
.	.	.
.	.	.
.	.	.
.	.	.

News articles.

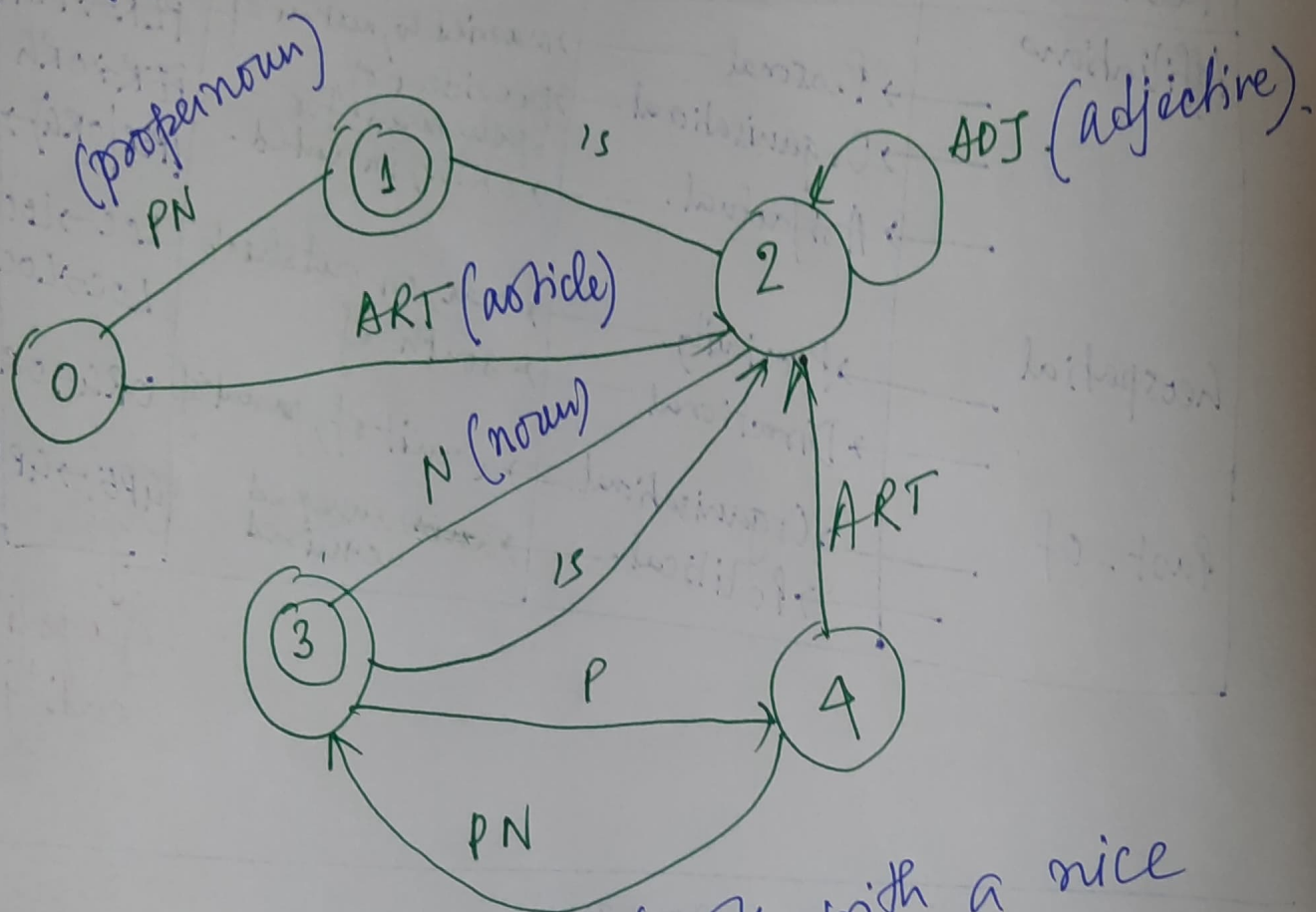
Relations.		Examples	Types.
Affiliations	→ Personal	→ married to, mother of	PER → PER
	→ Organisational	→ president of, spokesman for	PER → ORG
	→ Artificial.	→ owns, invented.	(PER ORG) → ART
Geospatial	→ Proximity	→ near, on outskirts	LOC → LOC
	→ Directional	→ south of	LOC → LOC
Part - of	→ Organisational	→ a unit of, parent of	ORG → ORG
	→ Political.	→ annexed annexed	GPE → GPE

↓
geopolitical entity.

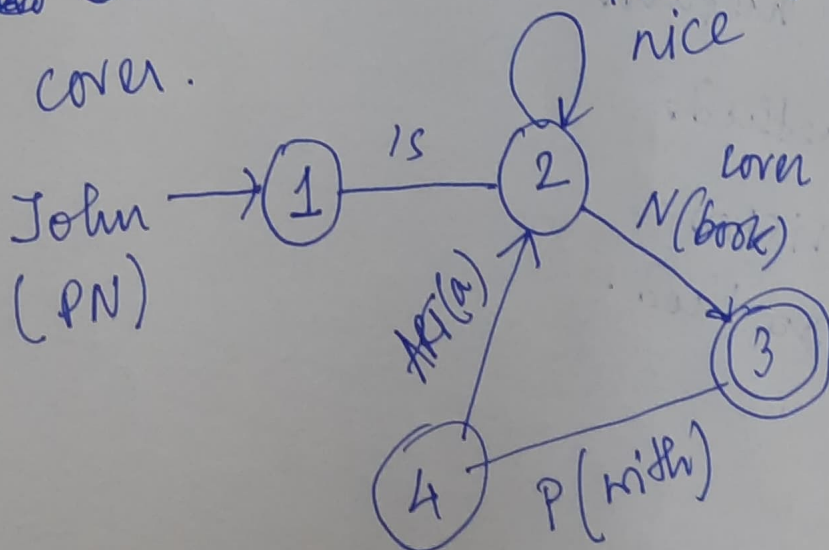
- Hand written rules (RELEX)
- Bootstrapping methods.
- Supervised methods
- Distant Supervision
- Unsupervised method.

RELEX based IE.

Use finite automaton for noun groups:



~~John's~~ John's interesting book with a nice cover.



many such
RELEX to
extract
info.

→ Determine which person holds what position in what organization.

[person], [position] of [org] ←

Vuk Draskovic, (leader) of the
Serbian Renewal Movement

[org] (named, appointed, etc.) [person] PREP [office].
(NATO) appointed Wesley Clark as Commander in Chief.
org person preposition. office.

→ Determine where an organization is located

[org] in [loc].

NATO headquarters in Brussels.

[org], [loc] (division, branch, headquarters, etc.)

KFOR, Kosovo headquarters.

Agar is a substance prepared from a mixture of red algae such as Gelidium for lab or industry use.