

Seminar: Text Mining for Historical Documents

YAGO – A Large Ontology from Wikipedia and WordNet

Weijia Shao

Knowledge? Text?



Which other singers were born when Elvis was born?

[Elvis Presley - Wikipedia, the free encyclopedia](#)

Elvis Presley was **born** on January 8, 1935 at around 4:35 a.m. in a two-room ... **Other singers** had been doing this for generations, but they **were** black. ...

en.wikipedia.org/wiki/Elvis_Presley [Cached](#) [Similar pages](#)

What can you learn?

Arrival of Air Force One in Peking,
02/21/1972



[Larger View](#)

[PDF/Print Version](#)

Air force one carrying the President of the United States

Arrival of Air Force One in Peking,
02/21/1972



[Larger View](#) [PDF/Print Version](#)

- Air Force One is the official air traffic control call sign of any United States Air Force aircraft carrying the President of the United States.

President of the United States from 1969–1974

Arrival of Air Force One in Peking,
02/21/1972



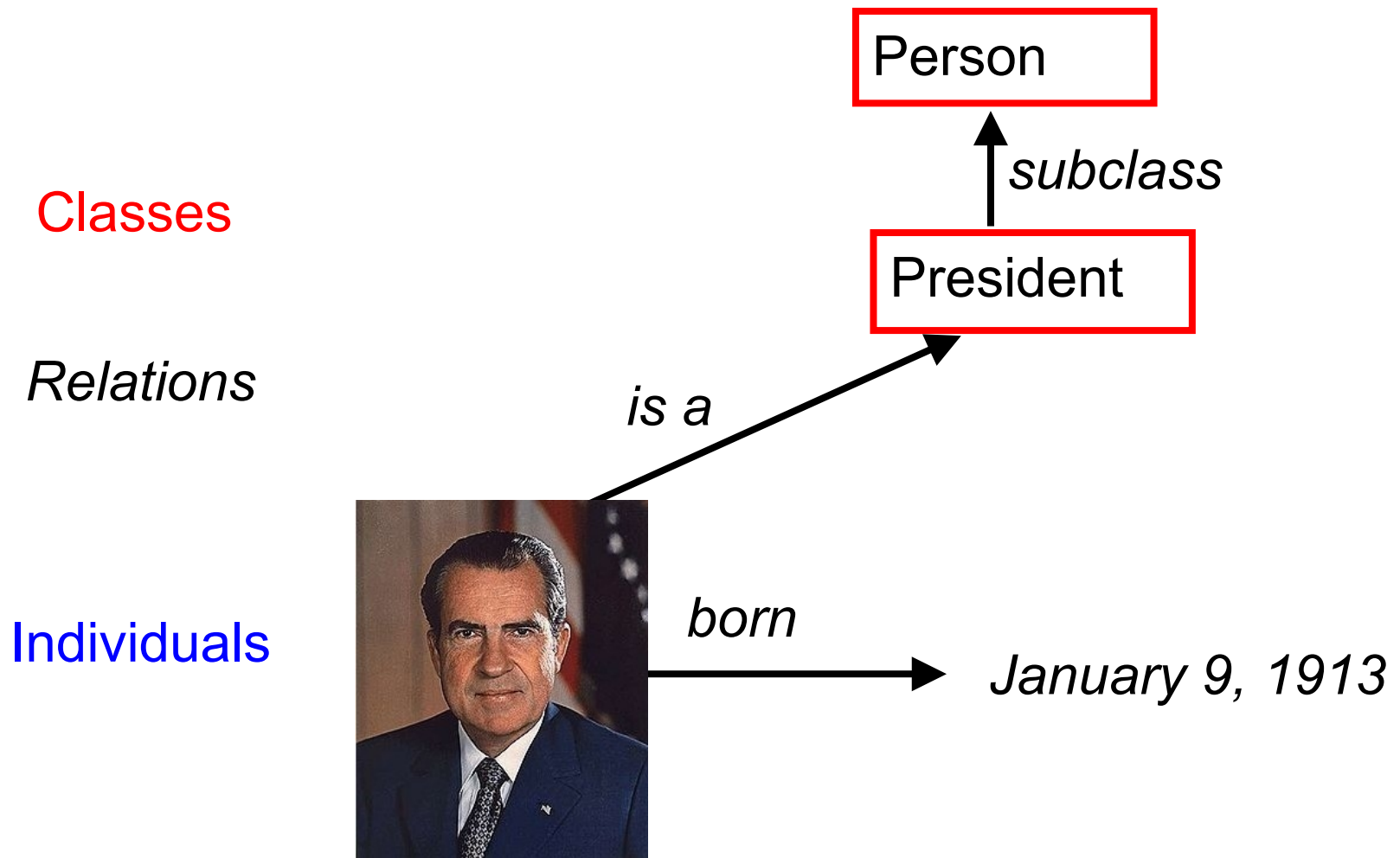
[Larger View](#) [PDF/Print Version](#)

- Air Force One is the official air traffic control call sign of any United States Air Force aircraft carrying the President of the United States.
- Richard Milhous Nixon (January 9, 1913 – April 22, 1994) was the 37th President of the United States from 1969–1974

Issues

- Knowledge representation
- Reasoning

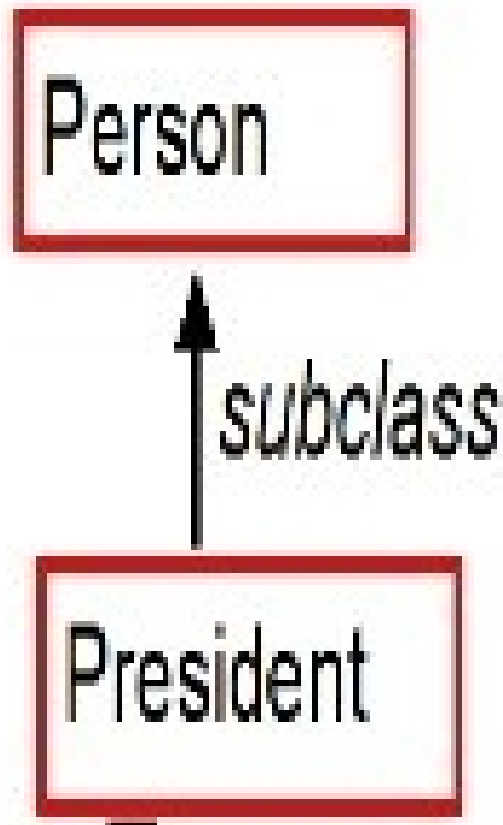
Solution: Ontology



Outline

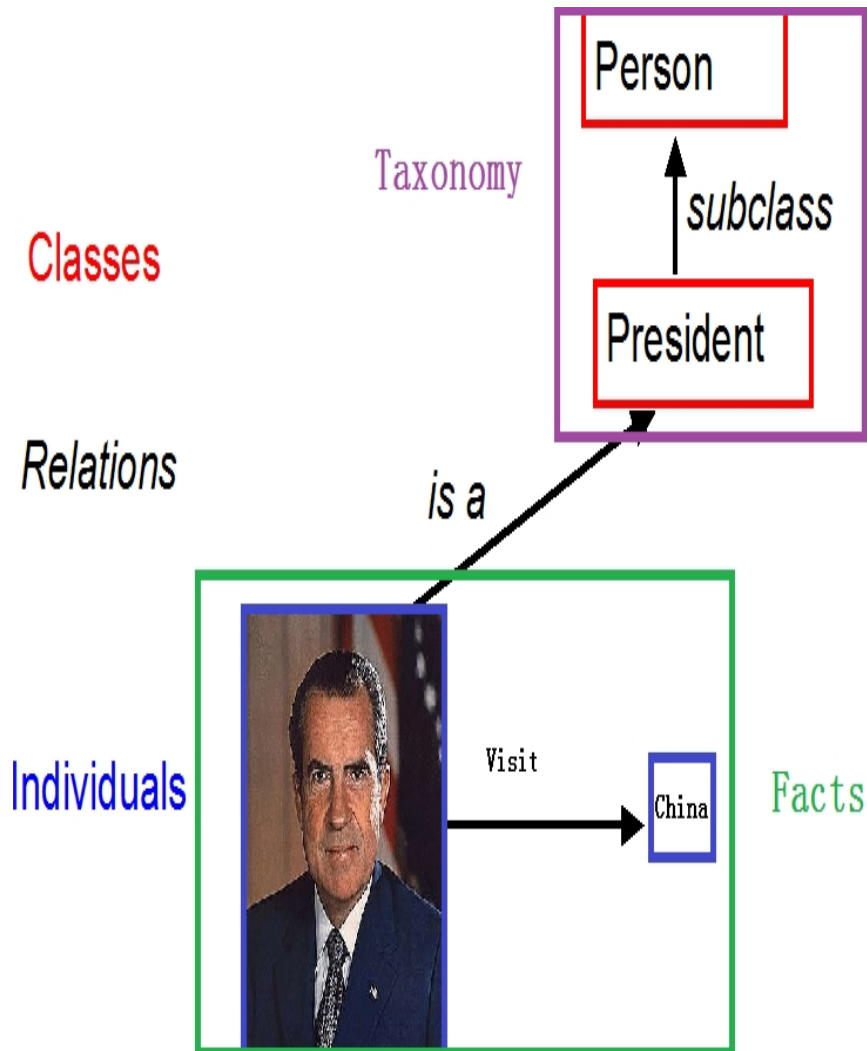
- Background
 - What is ontology?
 - What is YAGO?
- YAGO Model
 - How does it represent knowledge?
 - How does it reason?
 - Access knowledge?

What is ontology?



- A formal representation of
 - a set of classes
 - the relationships between those classes
- Used to reason

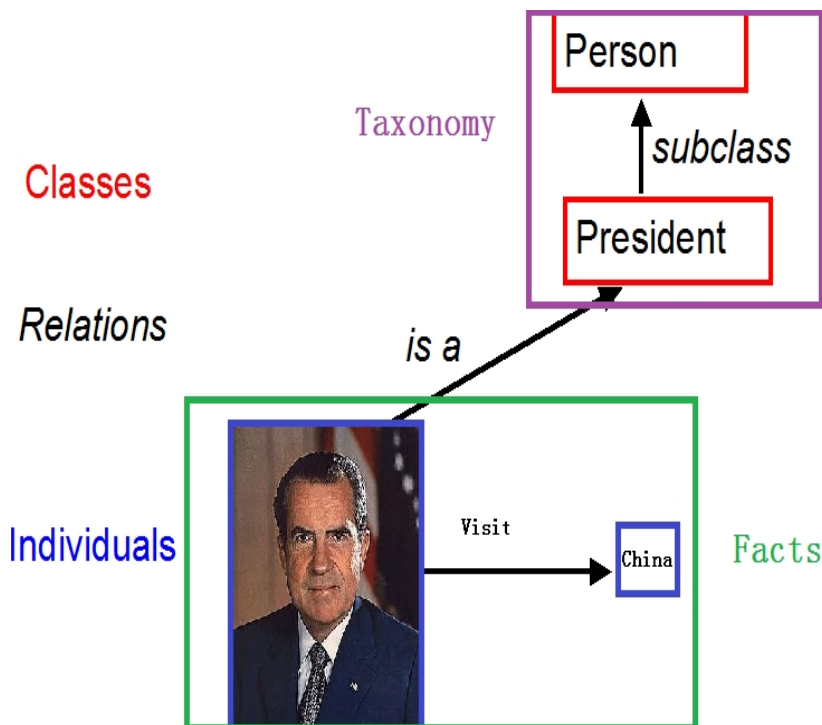
What is YAGO?



- A formal representation of
 - a set of classes and individuals
 - the relationships between them

What is YAGO?

- Facts are from category system and the infoboxes of Wikipedia



Categories: Richard Nixon | American anti-communists | American lawyers | Vietnam War | American Quakers | California Republicans | Deaths from stroke | States (1964–1980) | Members of the United States House of Representatives | American presidential pardons | Republican Party (United States) presidential Time magazine Persons of the Year | United States Navy officers | United States 1968 | United States presidential candidates, 1972 | United States vice-presidential figures | Nixon administration personnel | Eisenhower Administration personnel | 1994 deaths | 20th-century vice presidents of the United States | 20th-century | Vice Presidents of the United States

37th President of the United States

In office

January 20, 1969 – August 9, 1974

Vice President Spiro Agnew (1969–1973)

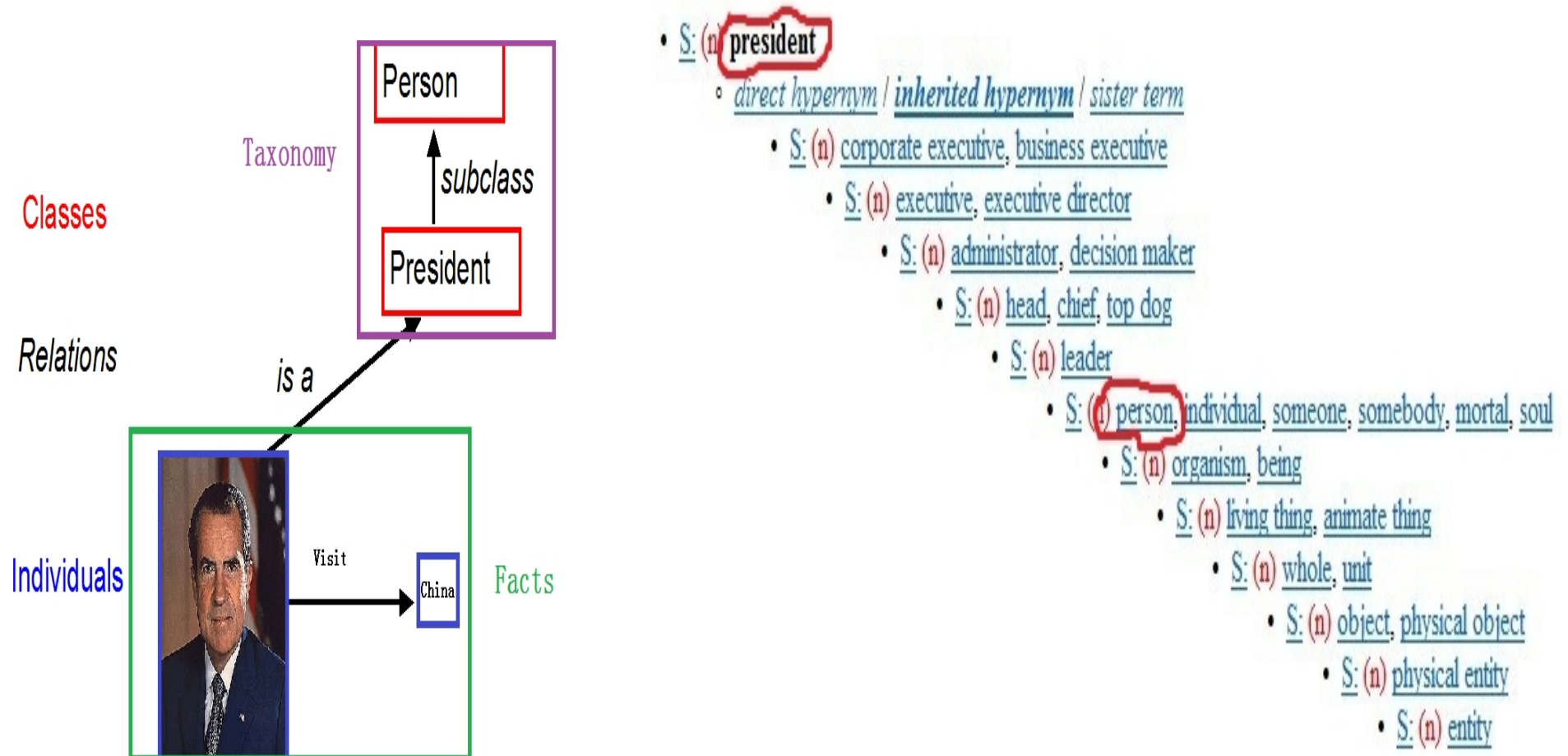
Gerald Ford (1973–1974)

Preceded by Lyndon B. Johnson

Succeeded by Gerald Ford

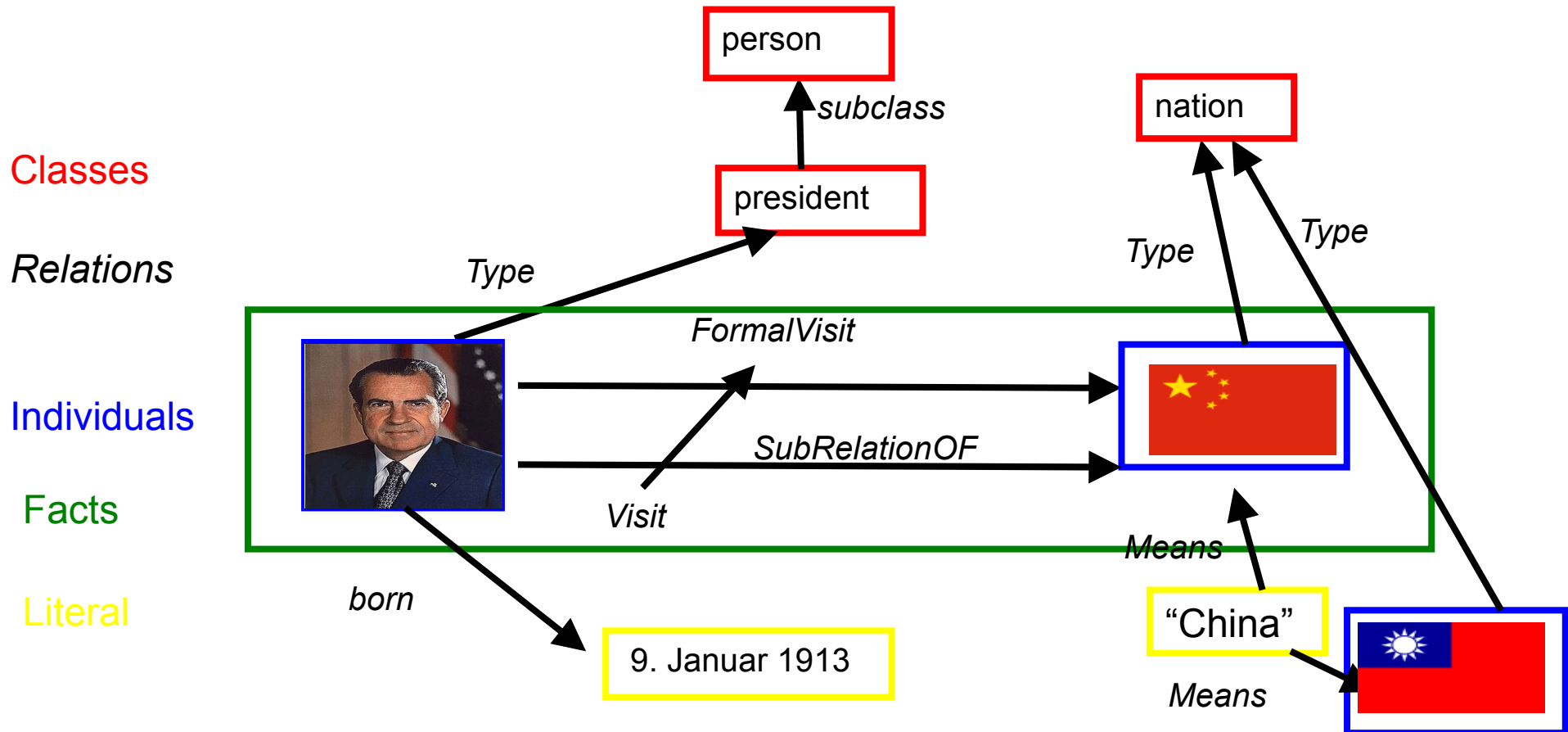
What is YAGO?

- Taxonomic relations are from WordNet



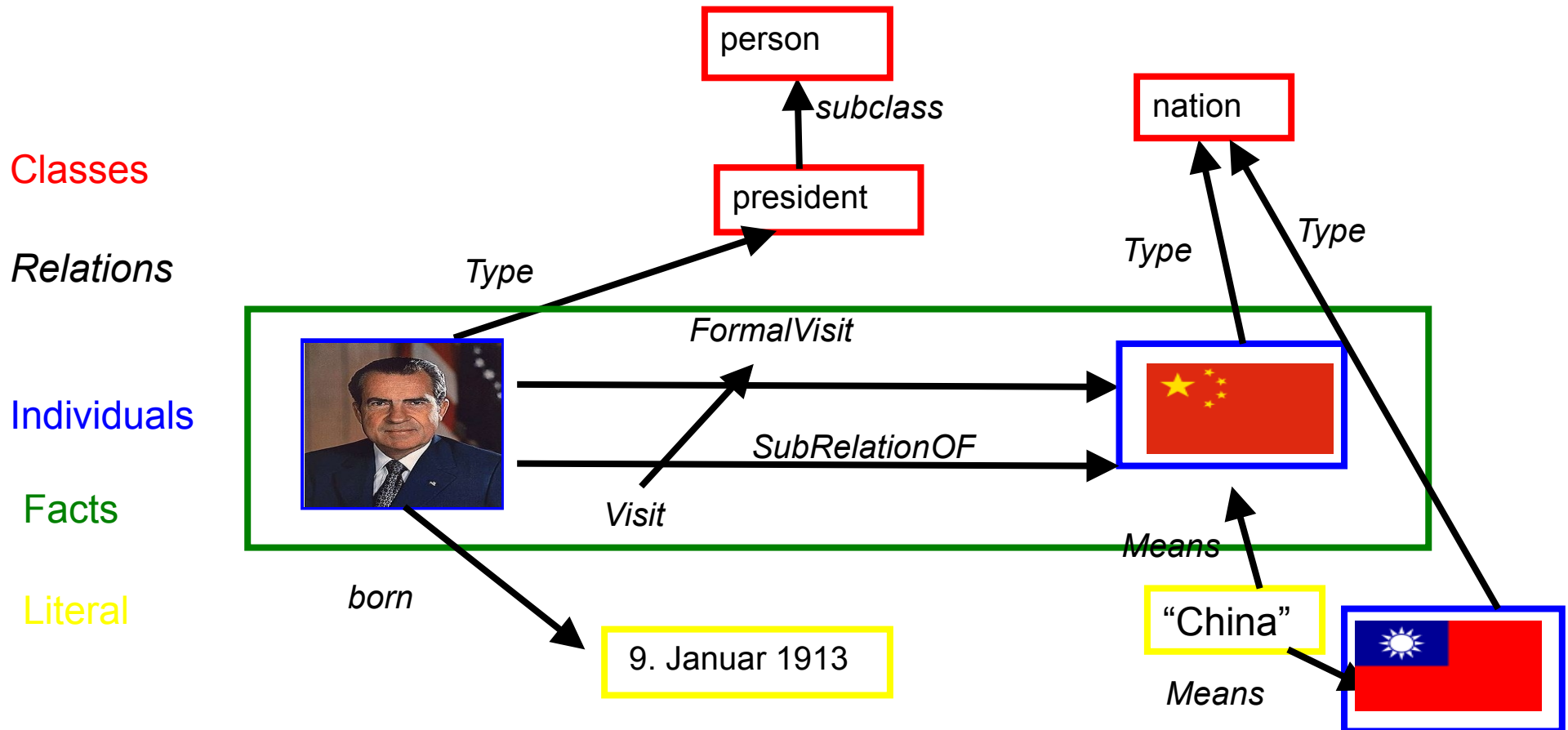
Yago Model

- Entity
 - All objects are entities. (classes, individuals, relations, literal,facts.....)



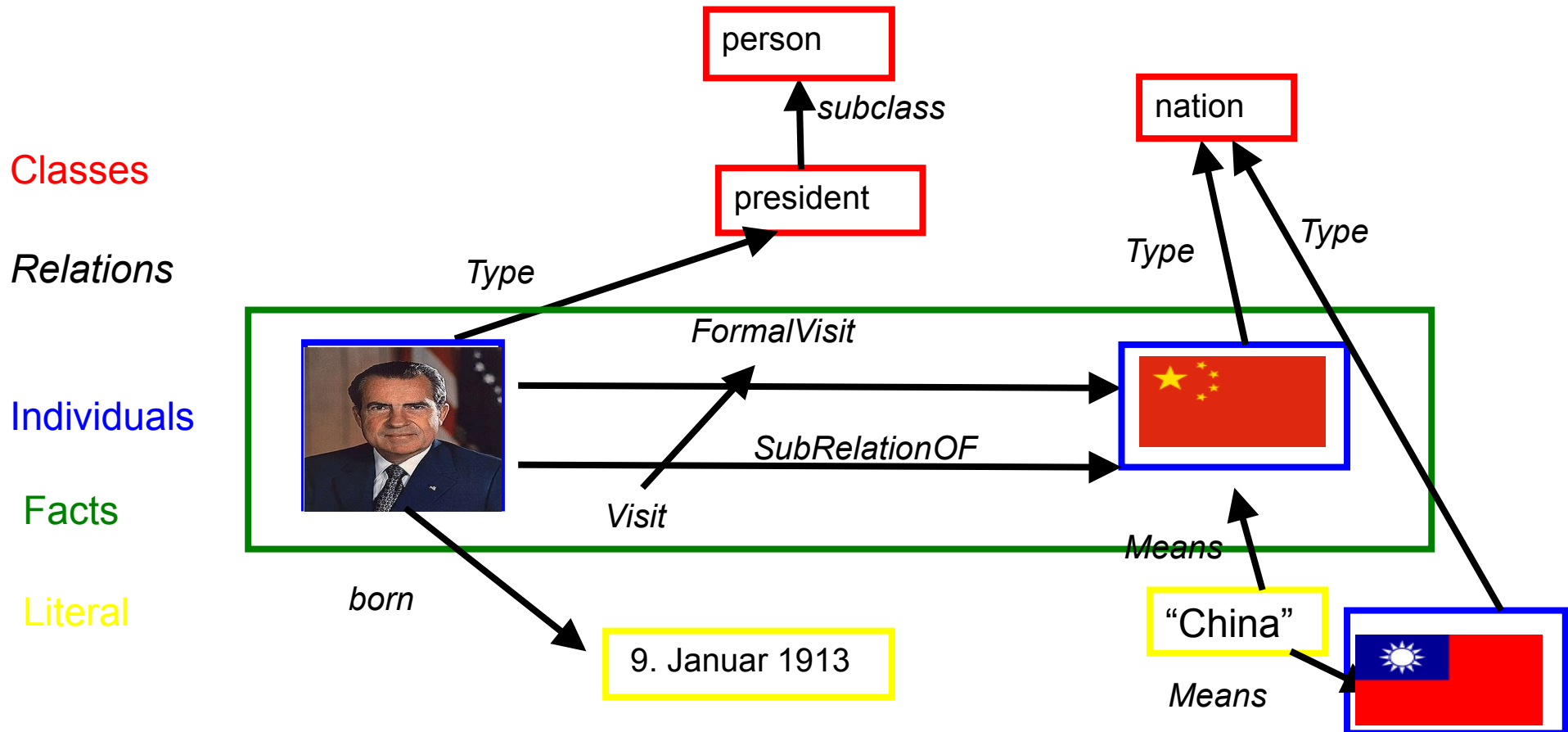
Yago Model

- Relation
 - Two entities can stand in a relation



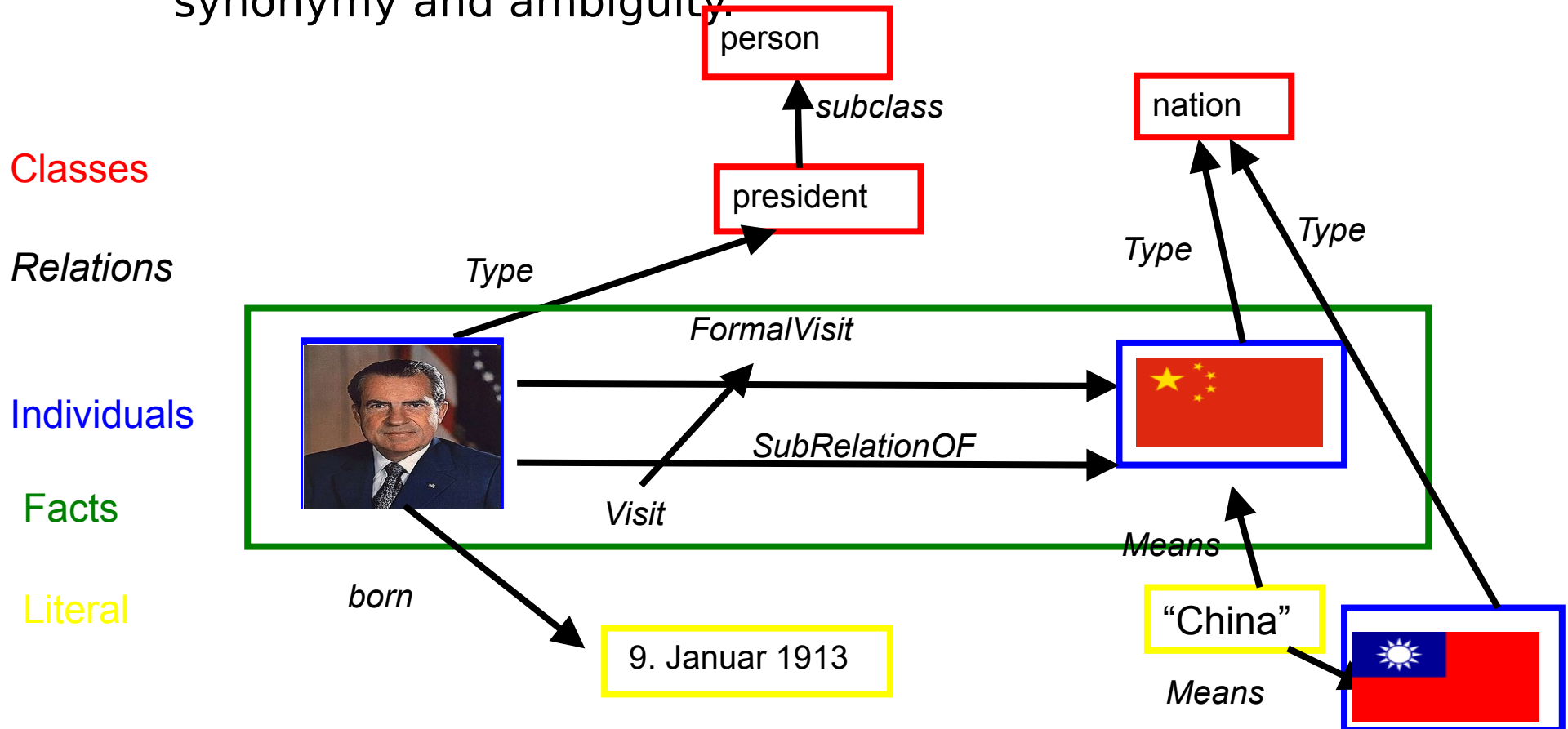
Yago Model

- Class: Concept, type....
- Individuals: instance of class



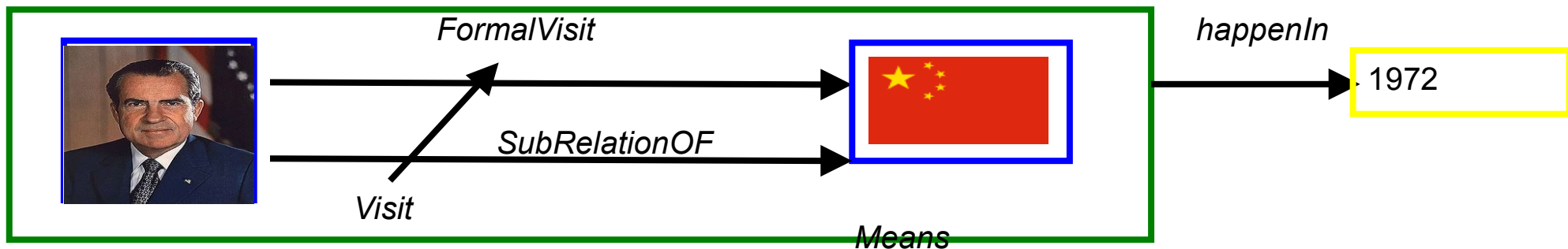
Yago Model

- Literal
 - Fixed value like date, word
 - One word refers more than one entities(word “china” in the documents about 2nd world war), use relation mean to deal with synonymy and ambiguity.



Yago Model

- Facts
 - Triple of (an entity, a relation, an entity)
 - Each fact has an ID
 - #1 Nixon(individual) visit(relation) China(individual)
 - Facts are entity
 - We can express n-ary relation: Nixon visit China in year 1972
 - #1(fact) happenIn(relation) 1972(individual)



Reasoning

- Reasoning =
 - Apply rules
 - To facts
 - Get more facts

Rules of YAGO

$$(1) \{(r_1, \text{SUBRELATIONOF}, r_2), (x, r_1, y)\} \hookrightarrow (x, r_2, y)$$

- Known fact:
 - FormalVisit is SubRelationOf Visit
 - Nixon FormalVisit China
- New fact
 - Nixon Visit China

Rules of YAGO

$$(2) \{(r, \text{TYPE}, \text{atr}), (x, r, y), (y, r, z)\} \hookrightarrow (x, r, z)$$

- Known fact:
 - LocatedIn Type atr
 - Saarbrücken LocatedIn Saarland
 - Saarland LocatedIn Germany
- New fact
 - Saarbrücken LocatedIn Germany

Rules of YAGO

(3) $\{(r, \text{DOMAIN}, c), (x, r, y)\} \hookrightarrow (x, \text{TYPE}, c)$

(4) $\{(r, \text{RANGE}, c), (x, r, y)\} \hookrightarrow (y, \text{TYPE}, c)$

- Known fact:
 - FormalVisit Domain Person
 - FormalVisit Range Nation
 - Nixon FormalVisit China
- New fact
 - Nixon Type Person
 - China Type Nation

Rules of YAGO

$$(5) \{(x, \text{TYPE}, c_1), (c_1, \text{SUBCLASSOF}, c_2)\} \hookrightarrow (x, \text{TYPE}, c_2)$$

- Known fact:
 - Nixon Type President
 - President SubClassOf Person
- New fact
 - Nixon Type Person

Query?

Query Form

YAGO-query:

?id0:	<input type="text" value="Albert_Einstein"/>	<input type="text" value="bornOnDate"/>	<input type="text" value="▼"/>	<input type="text" value="?EinsteinDate"/>
?id1:	<input type="text" value="?x"/>	<input type="text" value="bornOnDate"/>	<input type="text" value="▼"/>	<input type="text" value="?otherDate"/>
?id2:	<input type="text" value="?otherDate"/>	<input type="text" value="after"/>	<input type="text" value="▼"/>	<input type="text" value="?EinsteinDate"/>
?id3:	<input type="text" value="?x"/>	<input type="text" value="hasWonPrize"/>	<input type="text" value="▼"/>	<input type="text" value="Nobel_Prize"/>

?EinsteinDate = [1879-03-14](#)
?otherDate = [1898-09-24](#)
?x = [Howard Florey, Baron Florey](#)

?EinsteinDate = [1879-03-14](#)
?otherDate = [1947-02-04](#)
?x = [Dan Quayle](#)

Conclusion

- YAGO
 - Knowledge representation
 - Reasoning
 - Query
- Opens up new opportunities for historical documents mining
 - Historical documents = place + event + person
 - Subset of wikipedia

Thank you for attention!

Questions?