

Ontology of the Three Kingdoms

Natural Language Processing Project

by Roberto Salvaneschi

LUO GUANZHONG

The Romance of the Three Kingdoms

演

義

ABRIDGED EDITION

Aim of the Project

- ❖ Starting from the novel *Romance of the Three Kingdoms*, I wanted to design and develop an ontology about the historical period the book is set on.
- ❖ Dataset, used tools, ontology structure and techniques used to retrieve informations will be explained during the course of the presentation.

Dataset

- ❖ *Romance of the Three Kingdoms* is a classical Chinese novel where mostly real historical character interact with each other during the turbulent years between 165 AD to 280 AD.
- ❖ Even if there are some mystical elements, events are based on the *Records of the Three Kingdoms*, which is the Chinese official history text.

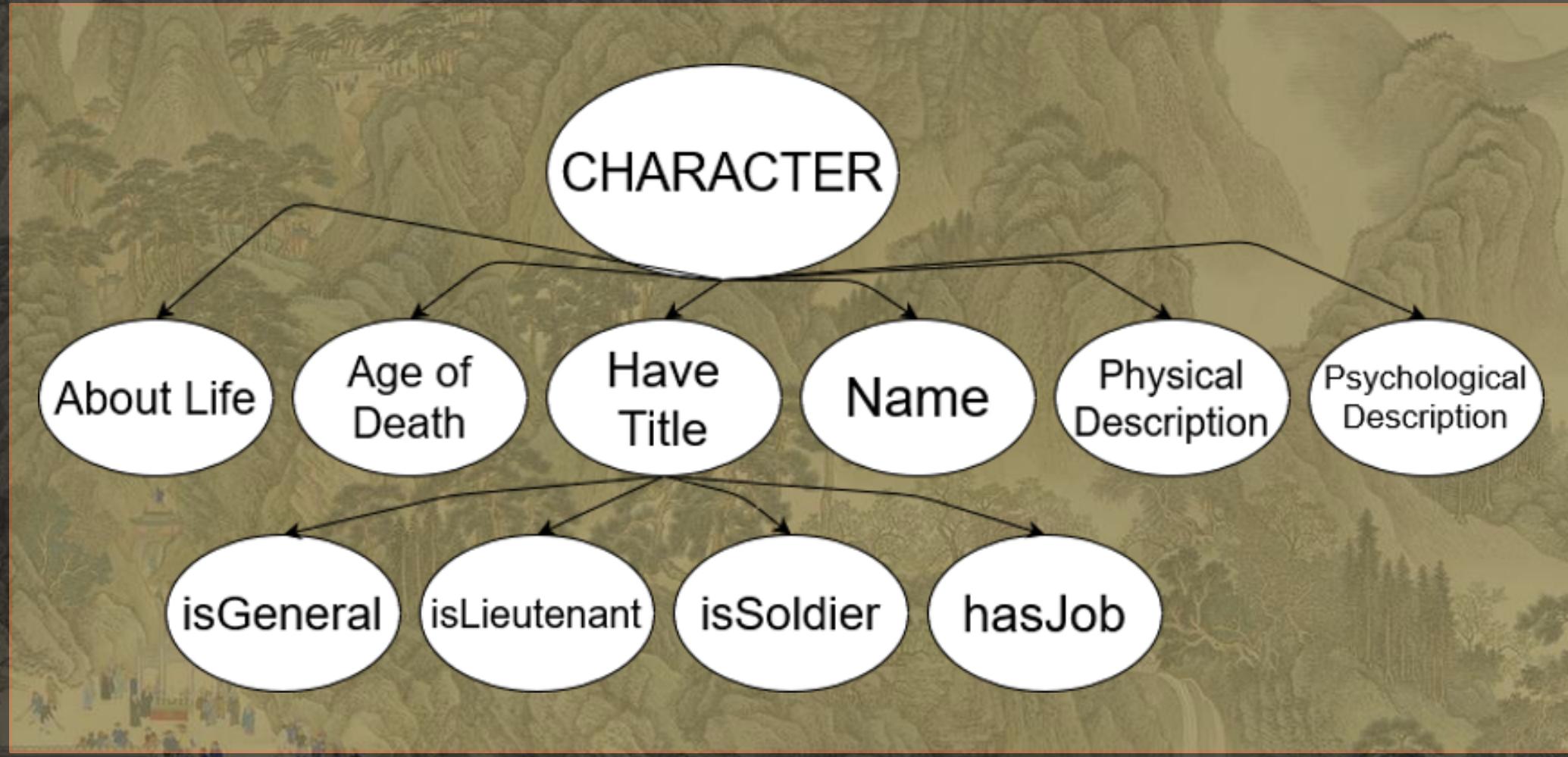
Used Tools

- ❖ *Protégé* has been used for building the ontology;
- ❖ *Python* has been used for developing the NLP-code together with the following libraries:
 - *nltk* for tokenizing words and gathering stop-words;
 - Standard libraries for string manipulation and system operations;
 - *owlready2* and *openpyxl* have been downloaded, but their usage didn't lead to any noteworthy results.

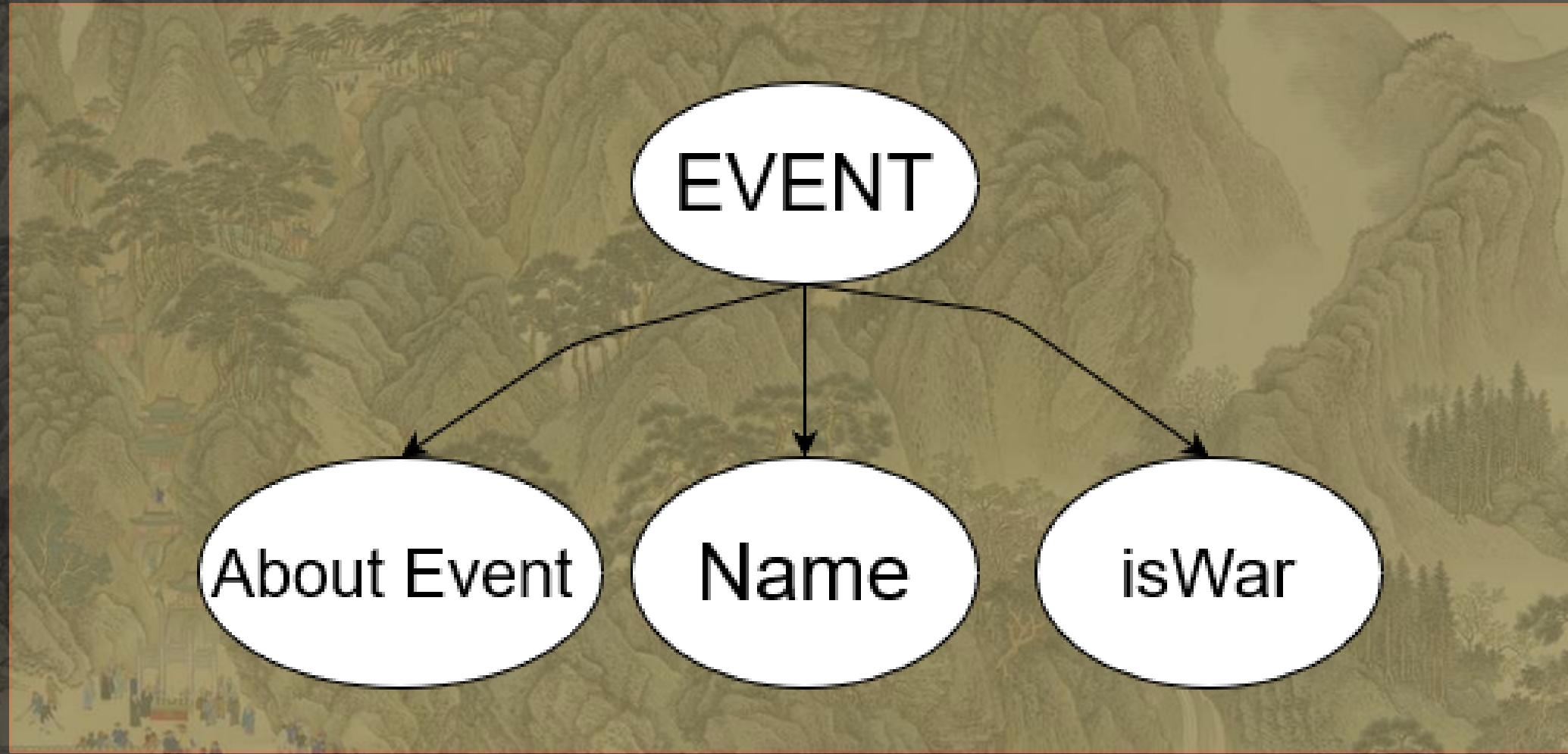
Ontology Structure

- ❖ Ontology is composed by five main classes representing most important concepts;
- ❖ Each class has its own data properties;
- ❖ Some classes may have some connections, those are represented with object properties.

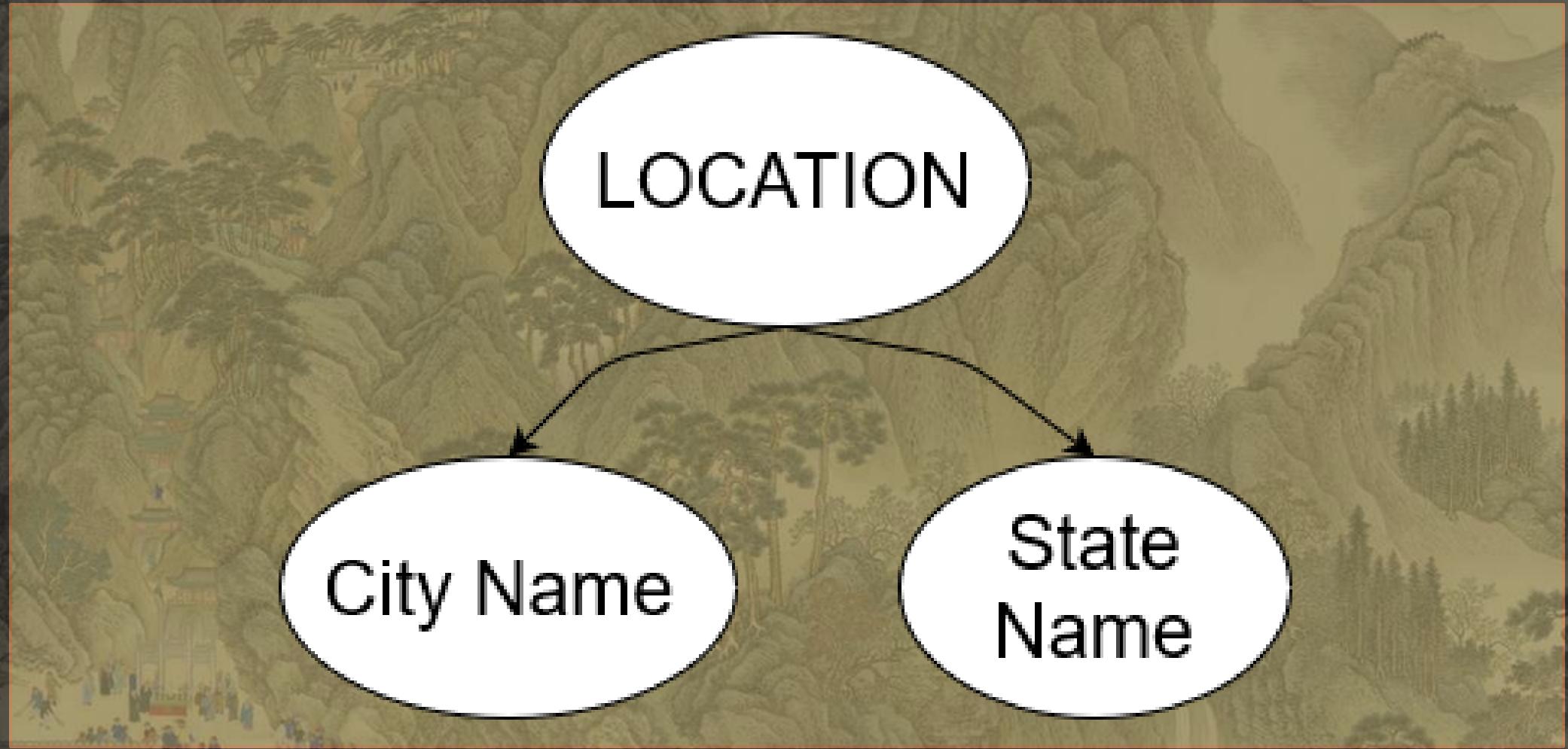
Class & Data Properties - Character



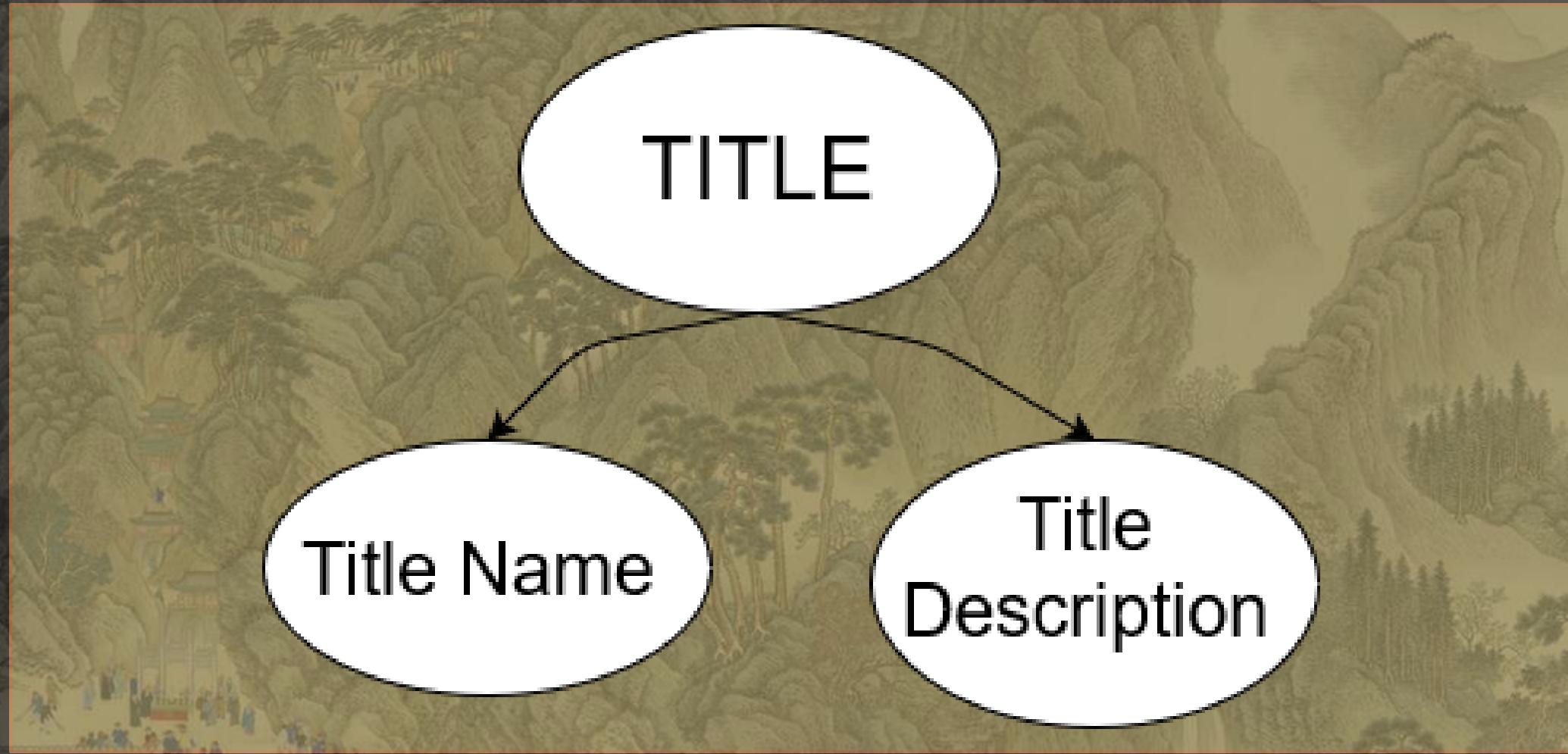
Class & Data Properties - Event



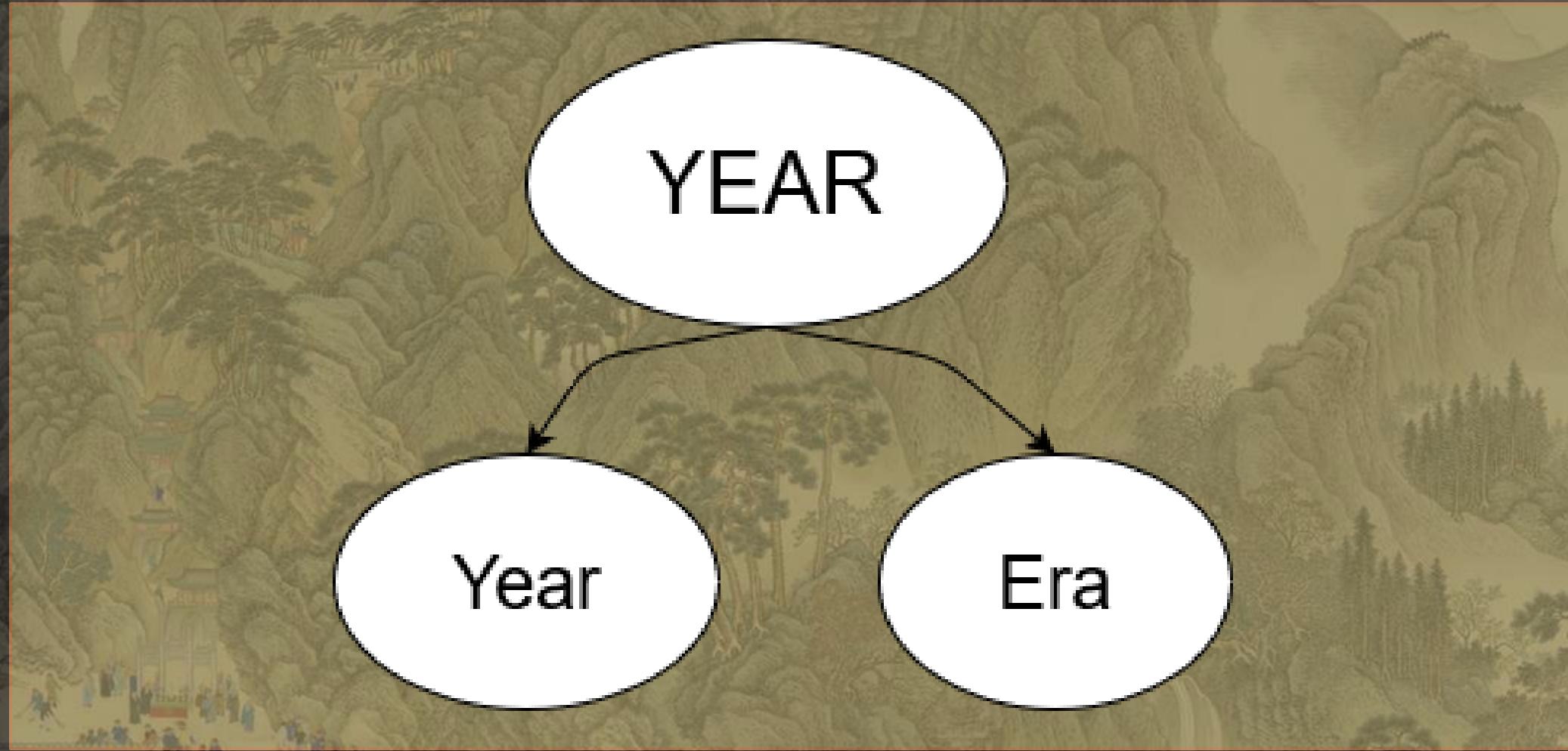
Class & Data Properties - Location



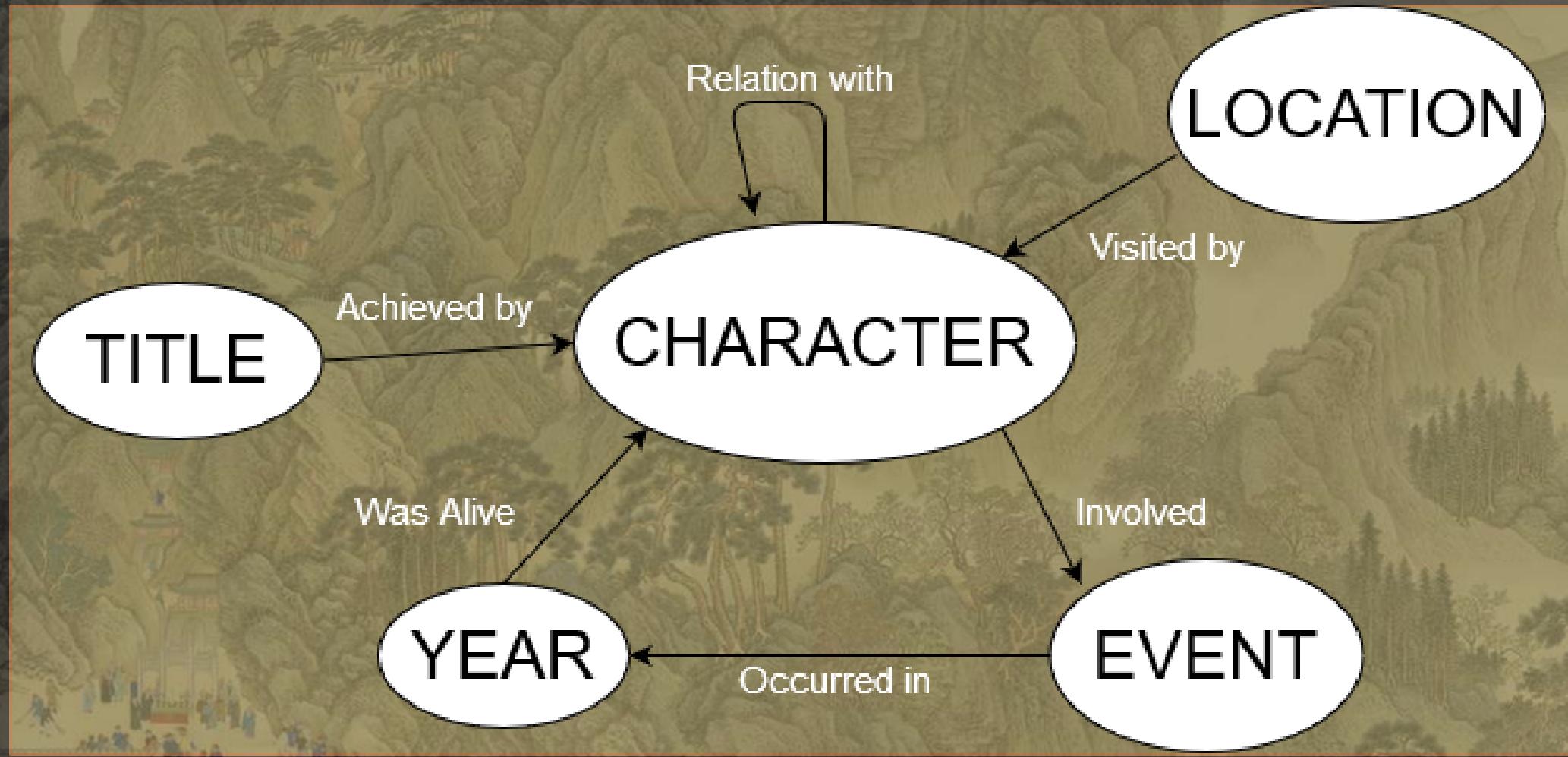
Class & Data Properties - Title



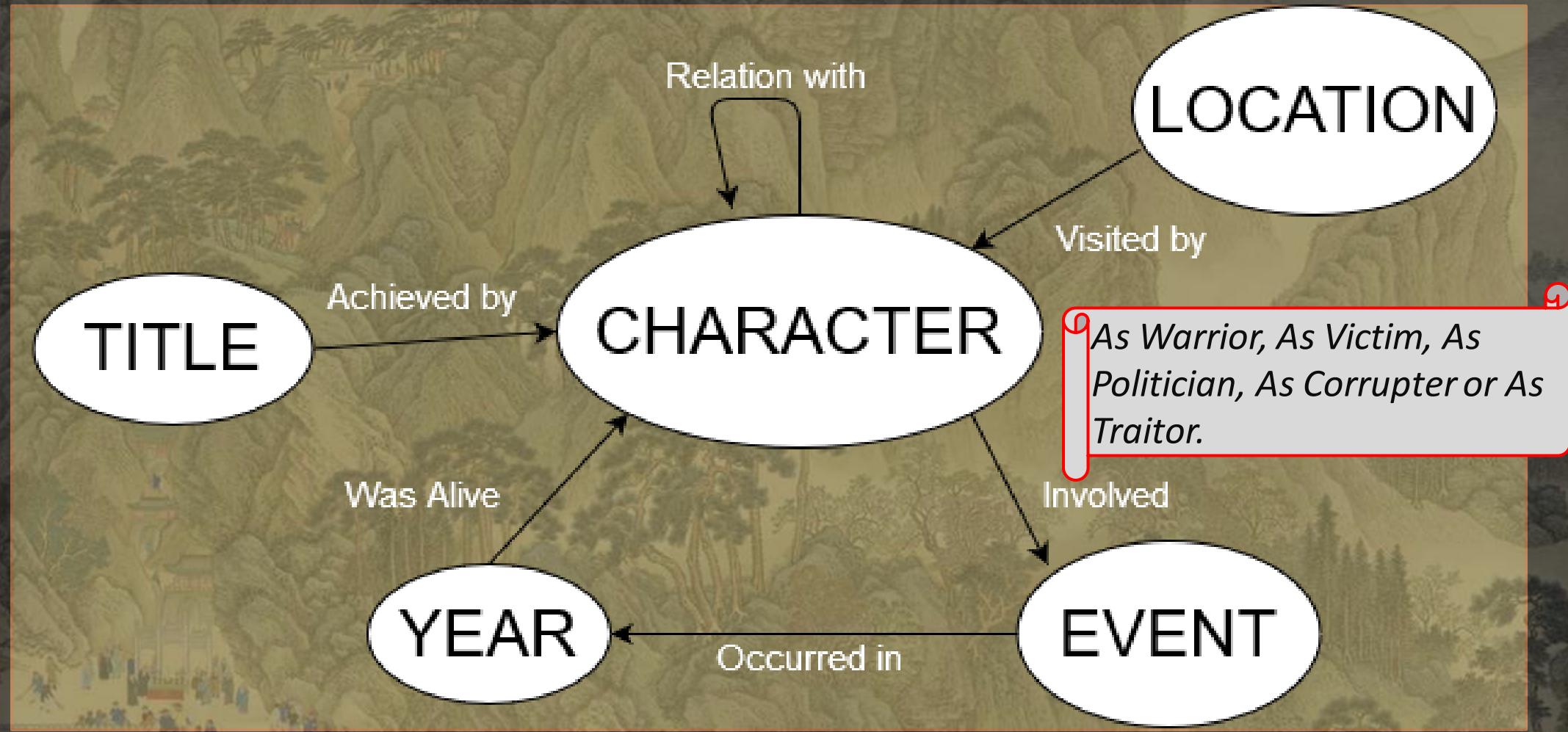
Class & Data Properties - Year



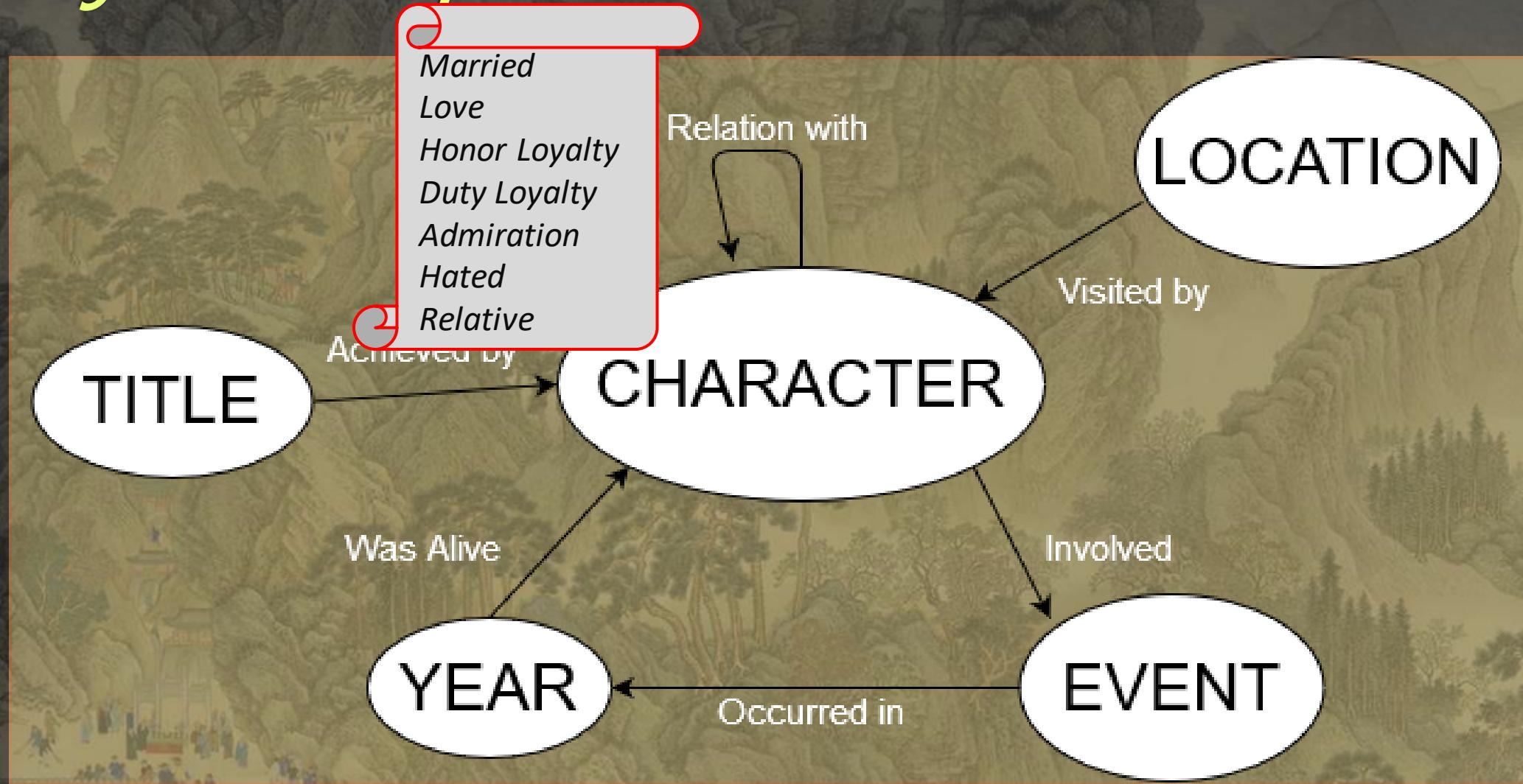
Object Properties



Object Properties



Object Properties



NLP Techniques with Python

- ❖ First I had to split the whole dataset in chapters:
 - *Reasoning on set of chapters made deductions simpler;*
 - *Python couldn't load all the book in one string.*
- ❖ Secondly I've split the text in sentences with the `split()` function, giving dot as input.

NLP Techniques with Python

- ❖ Once I had sentences, I've select all those where a certain Person appears, dividing them in blocks:
 - A block is made by the merging of the sentence where the searched Person appears, with the one before and one after.
- ❖ From those block, I've removed stop-words and computed Term Frequency.

NLP Techniques with Python

- ❖ By looking at most frequent words we can understand a lot about the selected character, but that was not enough to fill the ontology.
- ❖ Still, TF clearly highlighted which characters our Person have interact the most so I choose one of them and I've select all blocks where this second Person appears.

NLP Techniques with Python

- ❖ Recomputing TF on this new set may be an idea and repeating this procedure selecting meaningful words can also be useful, but in my case selected blocks weren't many so I've just read them all.
- ❖ From those sentences it was possible to understand relationships among characters and some other interesting informations useful for the ontology.

NLP Techniques with Python

- ❖ Starting from TF computed on a character is possible to determine what kind of role he/she played in the analysed events.
- ❖ This '*metric*' is calculated by combining TF with another list full of words commonly used by people related to the role we are interested in.

NLP Techniques with Python

- ❖ Creating those kind of lists may require time since you have to identify terms related to the role.
- ❖ In order to test this idea, I've naively created two lists:
 - *warWords.txt* to identify warriors;
 - *PoliWords.txt* to identify politicians.
- ❖ Both lists are full of anachronistic terms.

NLP Techniques with Python

- ❖ Results were surprisingly good giving the words used, even if greater information comed out of the ranking rather on the values obtained.
- ❖ I'm positive that with further testing and better lists, this technique could be a reliable choice.

Conclusions

- ❖ In general, the used approach proved to be good enough for retrieving information which can be useful to enrich the ontology.
- ❖ Of course with more time it could have been possible to better refine some elements and improve results precision and speedup the searching.



Thanks for the attention!