# Grafeno: Semantic Graph Extraction and Operation

Antonio F. G. Sevilla, Alberto Fernández-Isabel & Alberto Díaz

19 Sept. 2016

# Representation of textual information

Management and processing of information requires a good model for its description.

For text:

- Words (statistical approaches, n-grams)
- POS-tag (morphological analysis)
- Parse tree (syntactic analysis)

What about *meaning*?

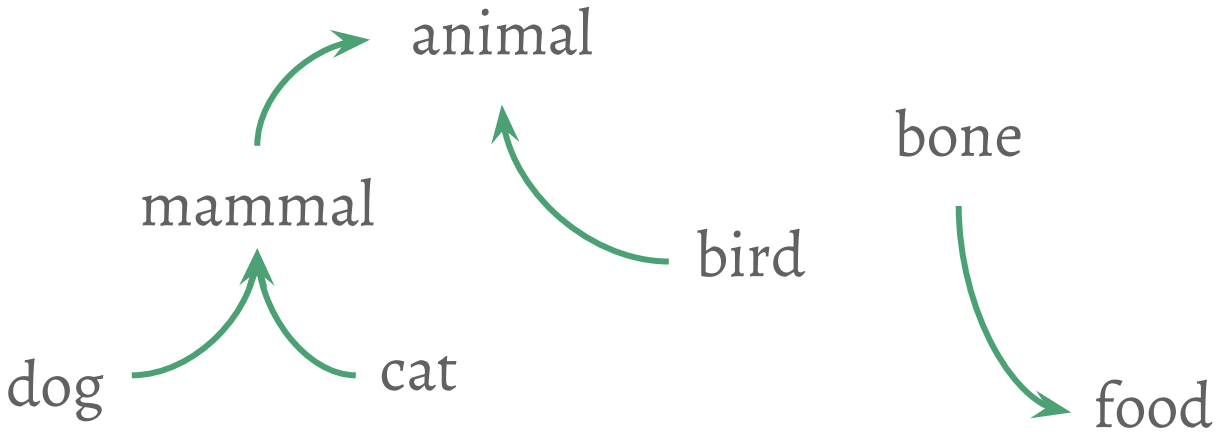
# Semantic background

- WordNet: synset
- Sowa: concept networks
- Prague FGD t-layer: (extended) dependency tree

# Our proposal: Semantic Graph

Nodes = concepts

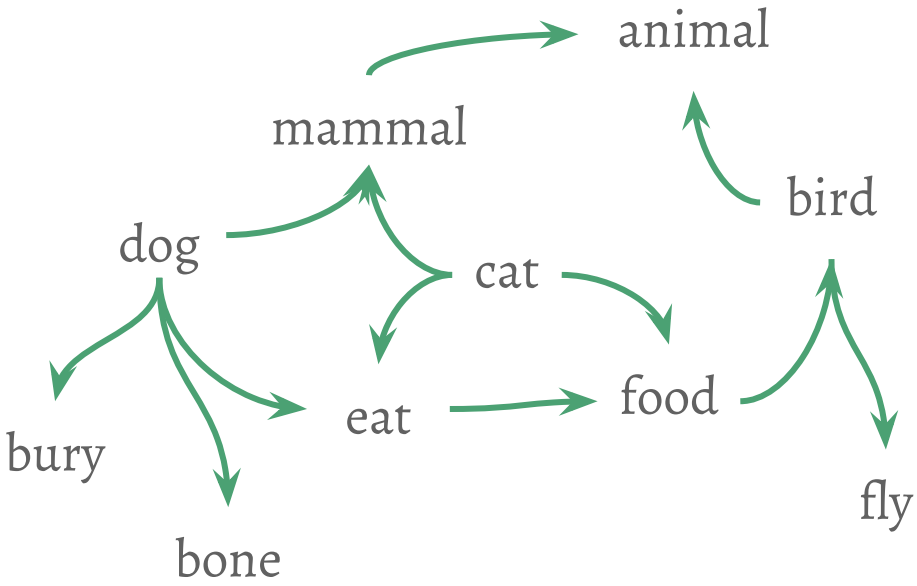
- ❏ Arbitrary identifiers
- ❏ Word lemmas
- ❏ Synsets



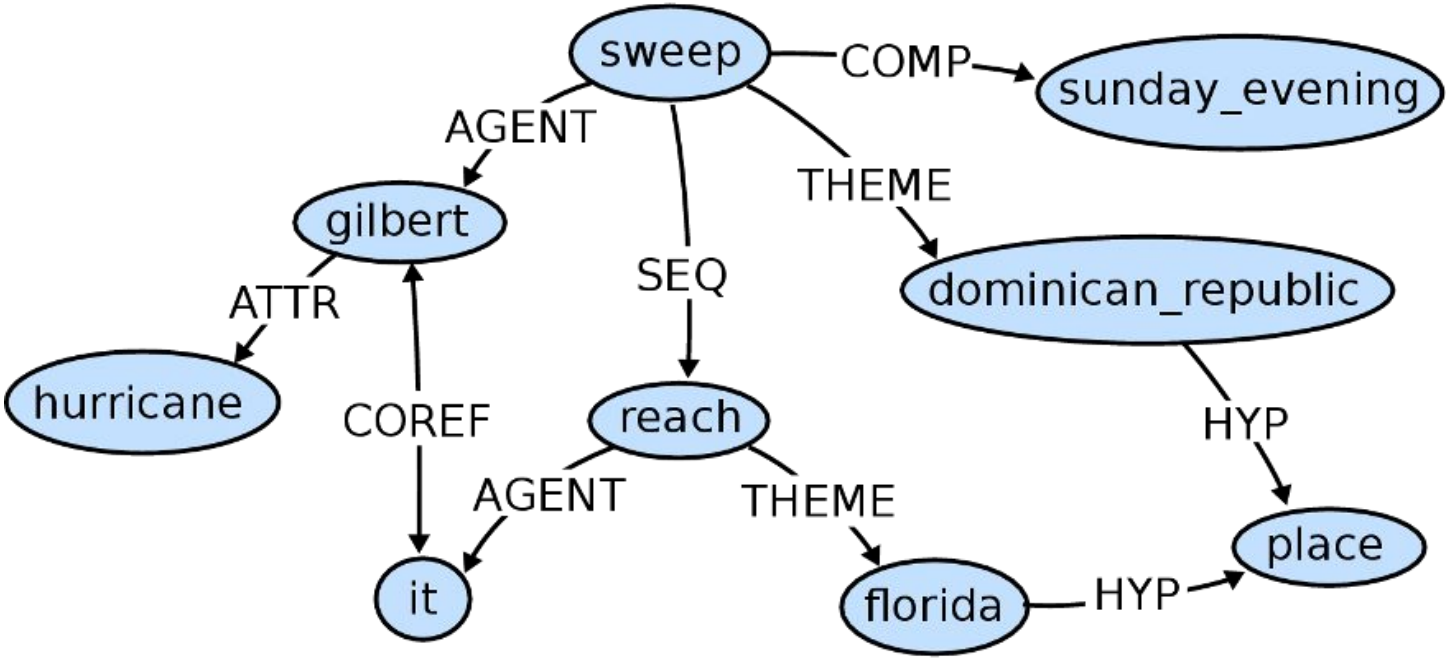
# Our proposal: Semantic Graph

Edges = semantic relations

- Grammatical
  (subject, object)
- Lexical
  (hypernym, similarity)
- Discourse
  (co-reference, conjunctive)

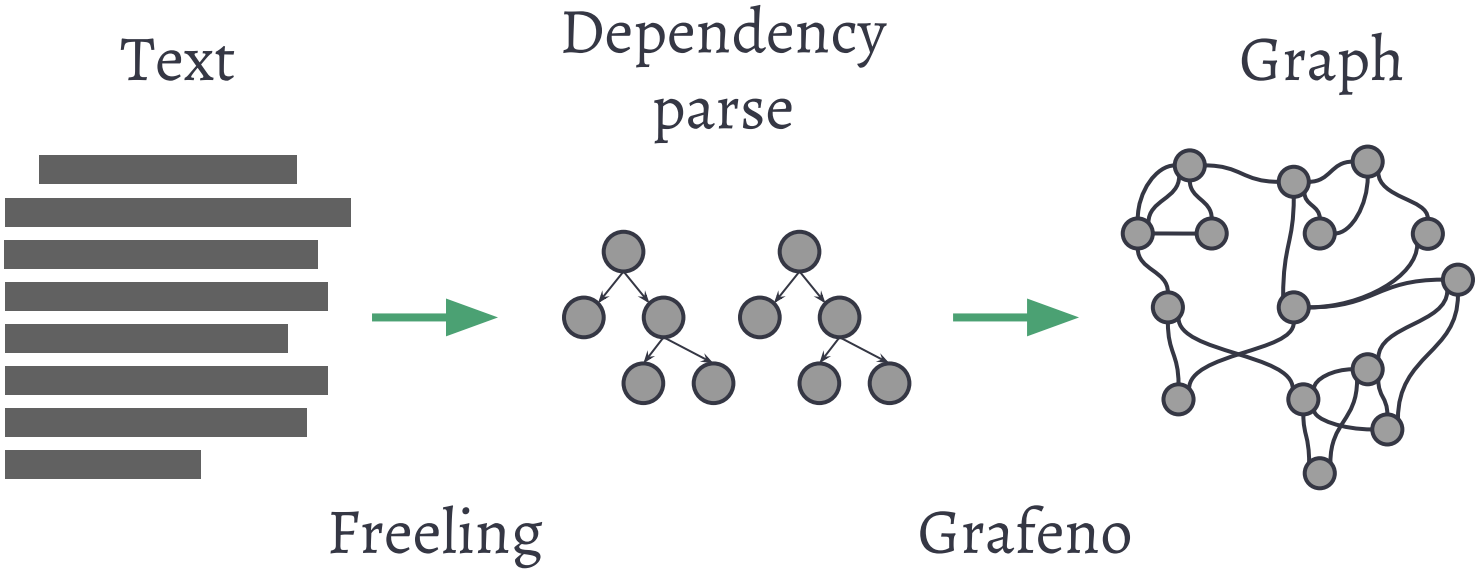


*“Hurricane Gilbert swept the Dominican republic Sunday evening. It then reached Florida.”*



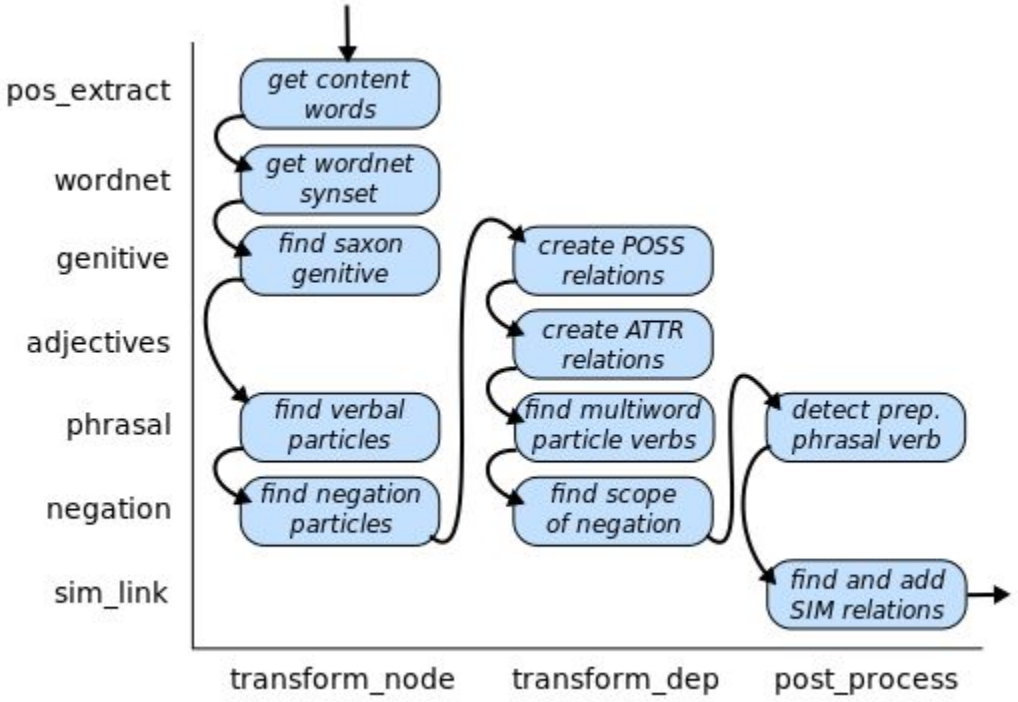
# **Grafeno**:
# Semantic Graph library

# Automatic graph extraction



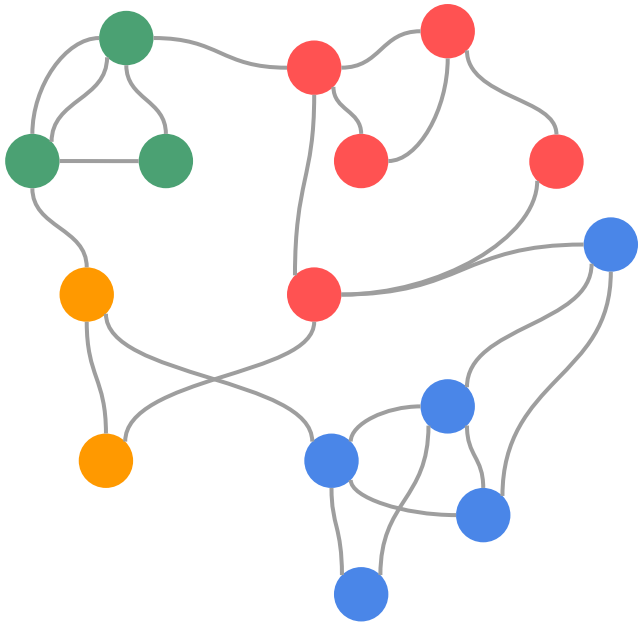
# Transformers

- Based on rules
- Expert knowledge
- External resources
- Modular



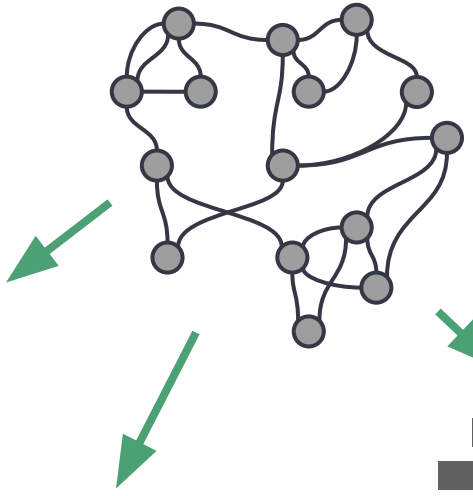
# Processing: operate

- Clustering
  (Degree-based, Markov, Louvain)
- Filtering
  (Edge renaming/deleting,
  domain extraction)
- Other
- Easy to extend



# Processing: linearize

```
Hurricane, Gilbert,
Sweep,
Dominican_Republic,
Sunday_evening...
```



```
<xml>
 <action>
  <verb>Sweep</verb>
  <agent>Gilbert</agent>
 </action>
...
```

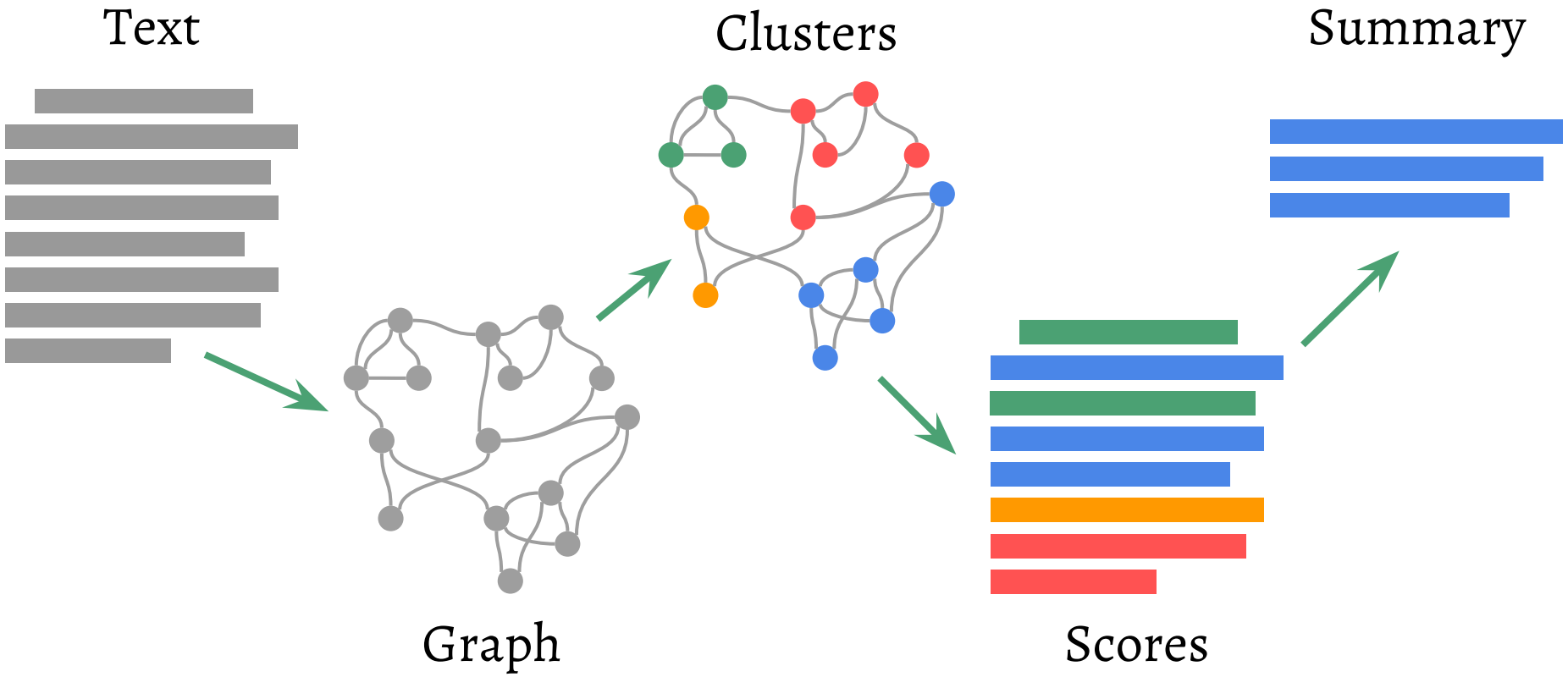
```
AGENT(Sweep,Gilbert)
THEME(Reach,Florida)
ATTR(Gilbert,Hurricane)
COREF(It_12,Gilbert)
```

# Pipeline scripts in YAML

```
%YAML 1.2
---
# Summarizes a text by extraction
transformers: [ pos_extract, sim_link,
     extend, unique, sentences ]
transformer_args:
     sempos: { noun: n }
     unique_gram: { hyper: [ True ] }
     extended_sentence_edges: [ HYP ]
operations:
     - op: cluster
       hubratio: 0.2
linearizers:
     - cluster_extract
linearizer_args:
     summary_length: 100
     summary_margin: 10
     normalize_sentence_scores: True
```

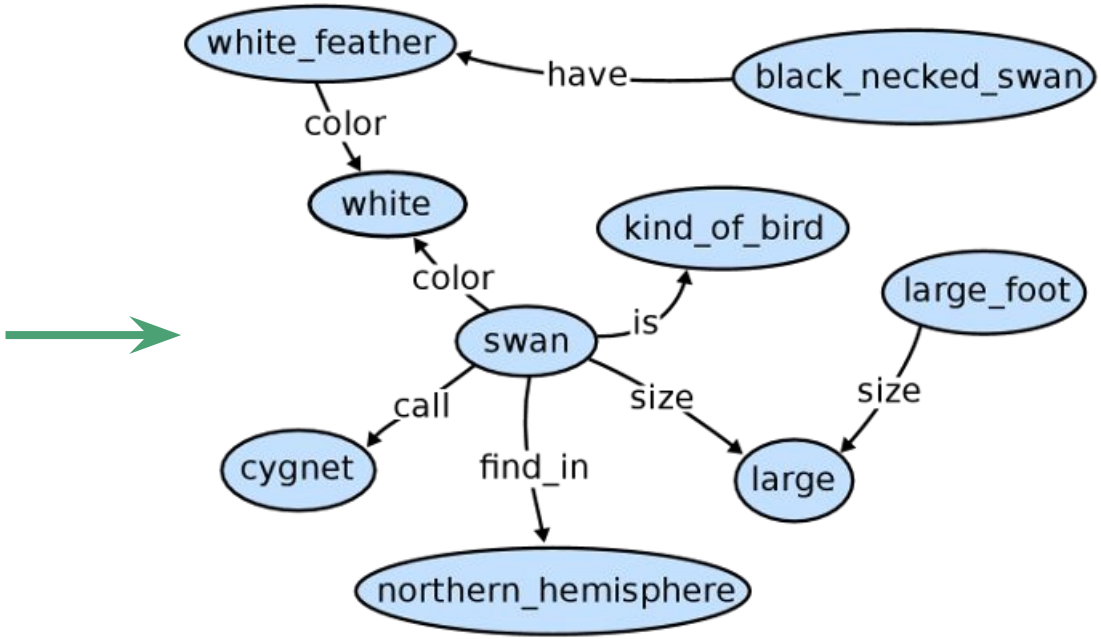
```
%YAML 1.2
---
# Extracts a concept map from a text.
transformers: [ pos_extract, wordnet,
     numerals, adjectives, negation,
     genitive, prepositions, attr_class,
     verb_collapse, specific_edges,
     unique, lenient ]
transformer_args:
     sempos: {      noun: n, adjective:
j }
     attach_adjectives: True
     keep_attached_adj: True
operations:
     - op: filter_edges
       remove: [ isa ]
       rename: { be: is }
       frequency: { max: 15, min: 0 }
     - op: spot_domain
linearizers: [ prolog ]
```

# Real example: summarization



# Real example: concept maps

Simple Wikipedia

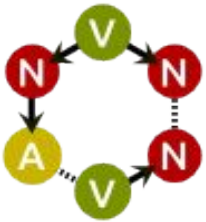


# Wrapping up

- ❏ Grafeno: a python library
- ❏ Extract semantic graphs from text
- ❏ Operate with them
- ❏ Code experiments in a script:
  - ❏ Execute it directly with CLI
  - ❏ Serve it over a web server

# Future Work

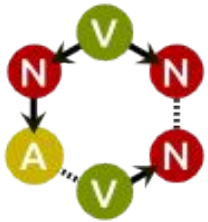
- ❏ Continue expanding linguistic rules
- ❏ Add support for more external tools
- ❏ Natural language generation




# END

ConCreTe European Project
http://conceptcreationtechnology.eu

NiL - Natural Interaction based on Language
http://nil.fdi.ucm.es




# END

Antonio F. G. Sevilla
afgs@ucm.es

Alberto Fernández-Isabel
afernandezisabel@ucm.es

Alberto Díaz
albertodiaz@fdi.ucm.es

Grafeno - semantic graph library
http://github.com/agarsev/grafeno