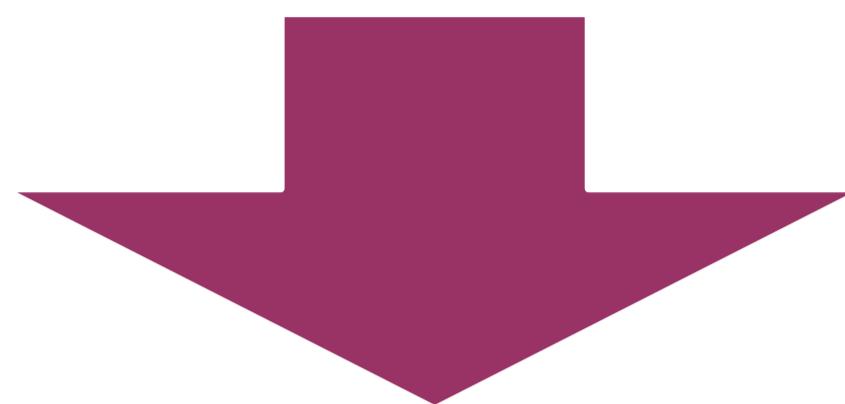


In a Nutshell: a 2-phase Frame-based Approach

Text

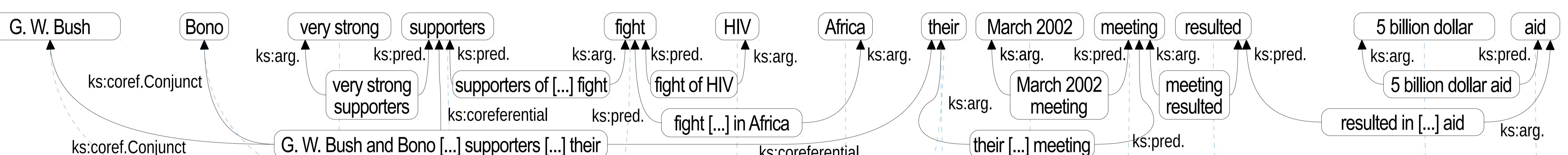
G. W. Bush and Bono are very strong supporters of the fight of HIV in Africa. Their March 2002 meeting resulted in a 5 billion dollar aid.



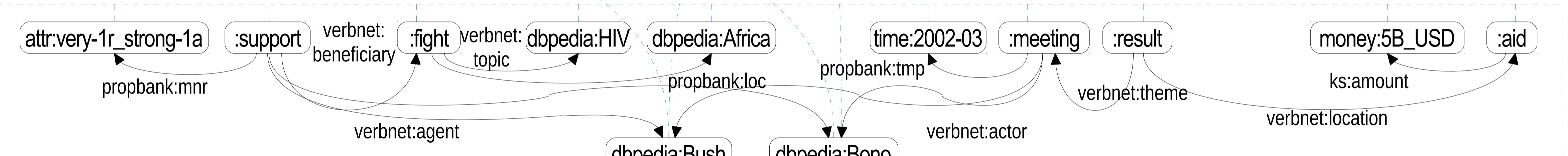
Phase 1 – Linguistic Feature Extraction

By performing several standard **NLP tasks**, a mention-based structured representation of the input text is built, organizing all the **annotations** produced by NLP tools (e.g., NERC, EL, TERN, SRL) in an **RDF graph of mentions** (i.e., spans of text denoting some entities or facts).

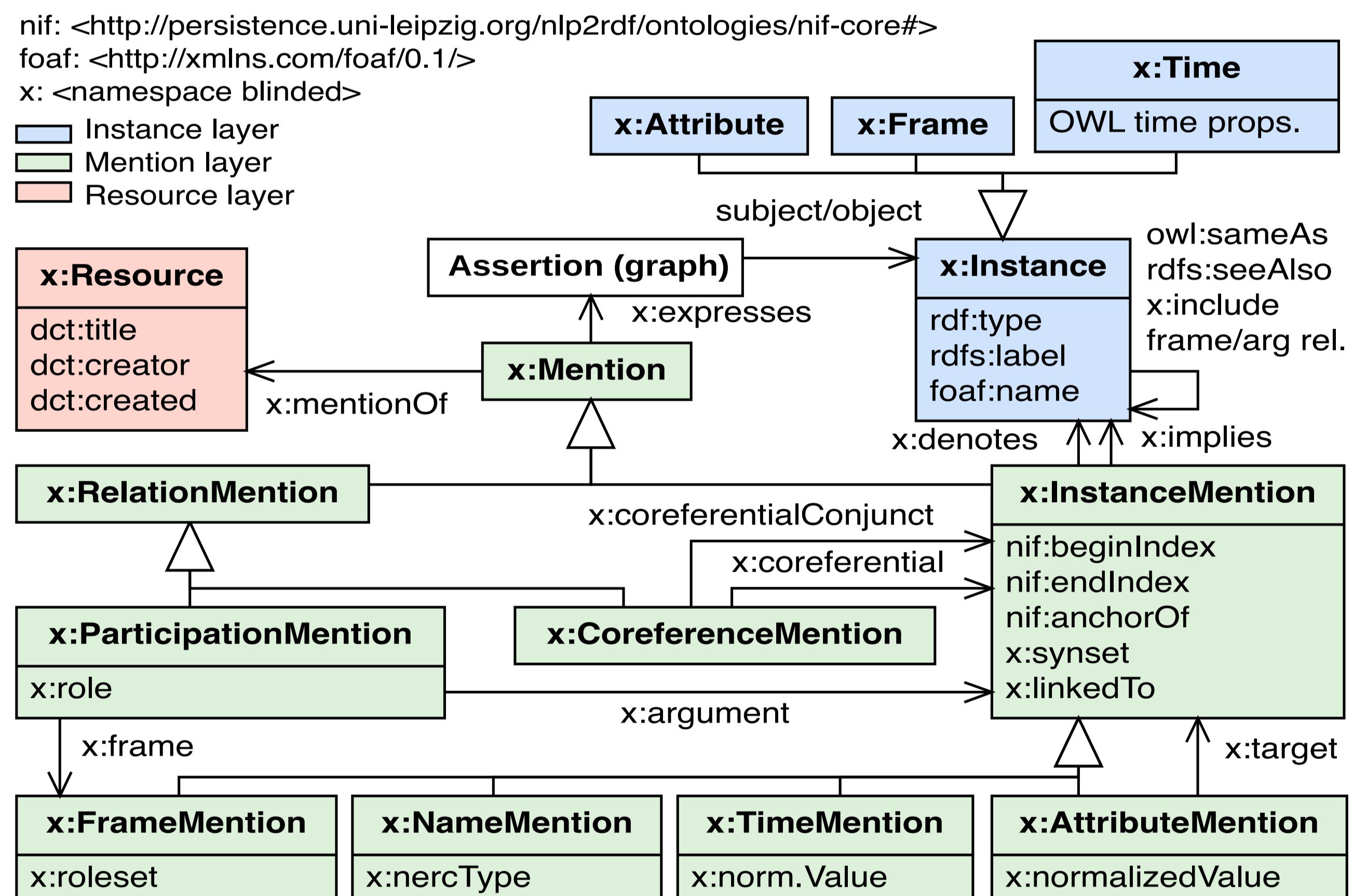
Mentions



Instances



RDF Data Model for Information Extraction



SPARQL-based Knowledge Distillation

Various types of **SPARQL rules**: instance creation, typing, naming, DBpedia linking, frame-role linking, coreference resolution

Example (Instance Creation for Argument Nominalization):

```
INSERT { ?m ks:denotes ?i ; ks:implies ?if ; ks:expresses ?g .  

GRAPH ?g { ?i a ks:Instance . ?if a ks:Instance , ks:Frame } }  

WHERE { ?m a ks:FrameMention ; nif:anchorOf ?a ; ks:roleset ?s .  

?s a ks:ArgumentNominalization . BIND (ks:mint(?m) AS ?g)  

BIND (ks:mint(concat(?a, " pred"), ?m) AS ?if)  

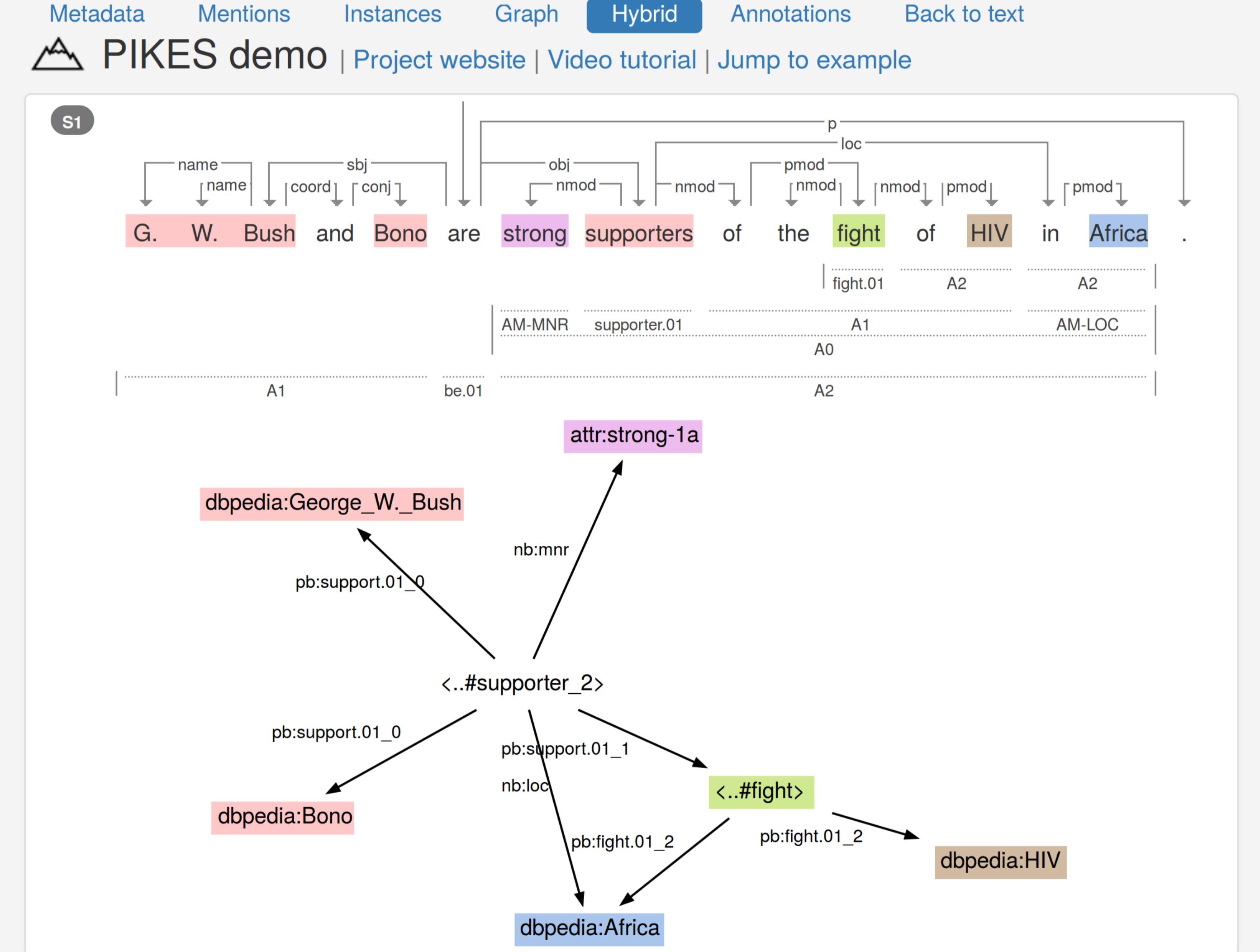
BIND (ks:mint(?a, ?m) AS ?i) }
```

Post-processing: inference, smushing, redundancy elimination, compaction.

References:

- Corcoglioniti, F., Rospocher, M., Palmero Aprosio, A.. *Extracting Knowledge from Text with PIKES*. In ISWC Posters & Demonstrations, 2015.
- Corcoglioniti, F., Rospocher, M., Cattoni, R., Magnini, B., Serafini, L., *The KnowledgeStore: a Storage Framework for Interlinking Unstructured and Structured Knowledge*. International Journal on Semantic Web and Information Systems, volume 11, 2015.

Graphical Rendering of Extracted Knowledge



Performances

Detecting and representing **frames** and **frame-role** relations:

→ precision: **0.716**
 → recall: **0.494**

Processing **large** document **corpora** (Simple English Wikipedia):

→ **110K pages** in about 507 core hours
 → processing-time **linearly scales** with the size of the text
 → **0.85 accuracy** in extracting triples about DBpedia entities

Powered By:

