

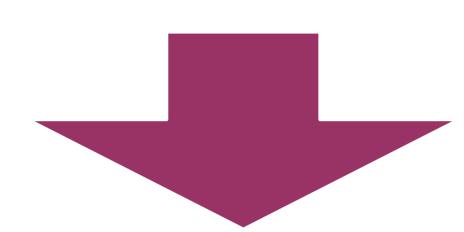
# PIKES

PIKES Is a Knowledge Extraction Suite pikes.fbk.eu



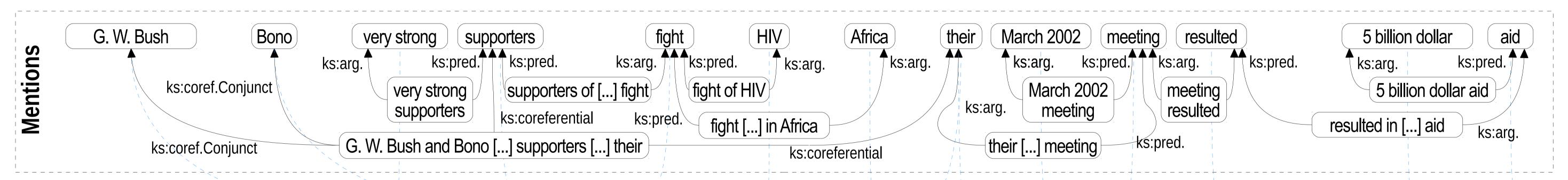
### In a Nutshell: a 2-phase Frame-based Approach

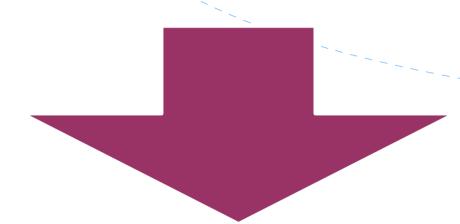
G. W. Bush and Bono are very strong supporters of the fight of HIV in Africa. Their March 2002 meeting resulted in a 5 billion dollar aid.



## **Phase 1 – Linguistic Feature Extraction**

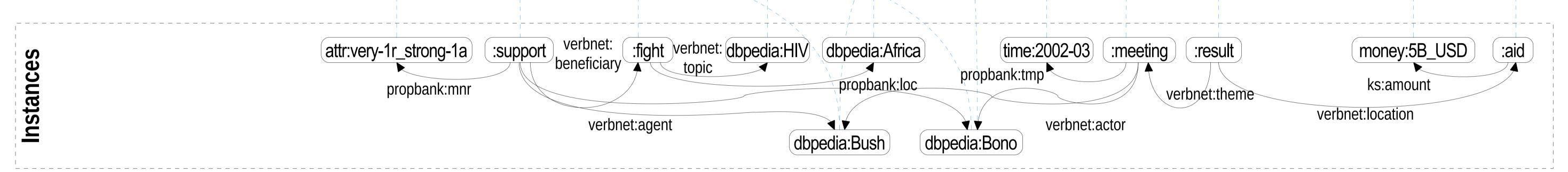
By performing several standard **NLP tasks**, a mention-based structured representation of the input text is built, organizing all the **annotations** produced by NLP tools (e.g., NERC, EL, TERN, SRL) in an **RDF graph of mentions** (i.e., spans of text denoting some entities or facts).





#### Phase 2 – Knowledge Distillation

The mention graph is processed via **SPARQL rules** to distill a **knowledge graph**, where each node uniquely identifies an **entity** of the world, **event** or **situation**, and arcs represent **relations** between them (e.g., the participation and role of an entity in an event).



#### **RDF Data Model for Information Extraction** nif: <a href="http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#">http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#</a> foaf: <a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/> x:Time x: <namespace blinded> OWL time props. x:Attribute x:Frame Instance layer Mention layer Resource layer subject/object owl:sameAs **Assertion (graph)** x:Instance rdfs:seeAlso x:Resource x:include x:expresses rdf:type dct:title frame/arg rel. rdfs:label dct:creator x:Mention foaf:name dct:created x:mentionOf x:implies x:denotes 个个 x:RelationMention x:InstanceMention x:coreferentialConjunct nif:beginIndex x:coreferential nif:endIndex nif:anchorOf x:ParticipationMention x:CoreferenceMention x:synset x:linkedTo x:role x:argument x:frame x:target x:FrameMention x:NameMention x:TimeMention x:AttributeMention x:nercType x:norm.Value x:normalizedValue x:roleset

#### 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):

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

#### Graphical Rendering of Extracted Knowledge Metadata Back to text PIKES demo | Project website | Video tutorial | Jump to example S1 [coord] [conj — nmod – [pmod] **Africa** and Bono are strong supporters fight.01 A2 AM-MNR supporter.01 AM-LOC be.01 **A**1 **A2** attr:strong-1a dbpedia:George\_W.\_Bush nb:mni pb:support.01 <..#supporter\_2> pb:support.01\_0 pb:support.01 nb:loc <..#fight> dbpedia:Bono pb:fight.01\_2 pb:fight.01\_2 dbpedia:HIV dbpedia:Africa

#### 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

#### 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.

#### Powered By:



