



MOOD-TC

a Multilingual Ontology Driven Text Classifier

Angelo Ferrando, Silvio Beux, Viviana Mascardi and Paolo Rosso

Università di Genova, Italy

Universitat Politècnica de València, Spain



THE PROBLEM



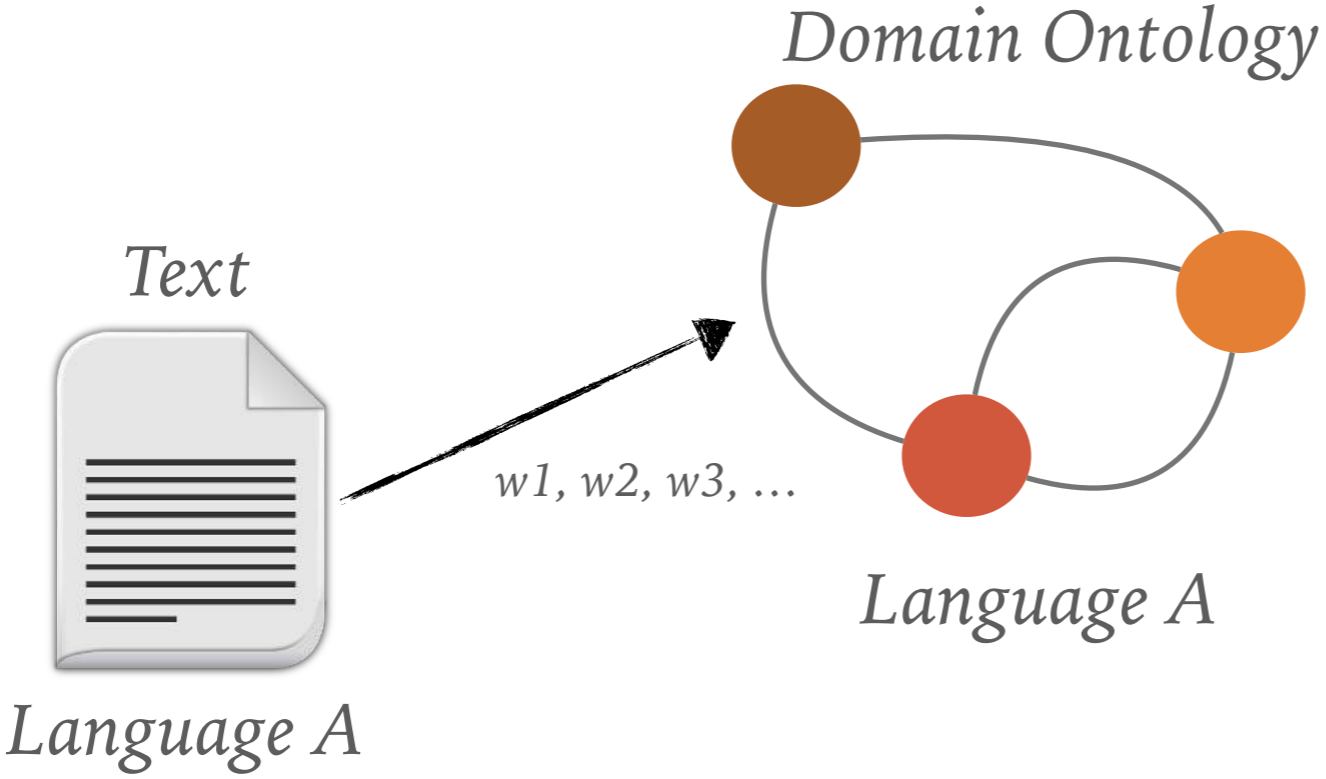
THE PROBLEM

Text

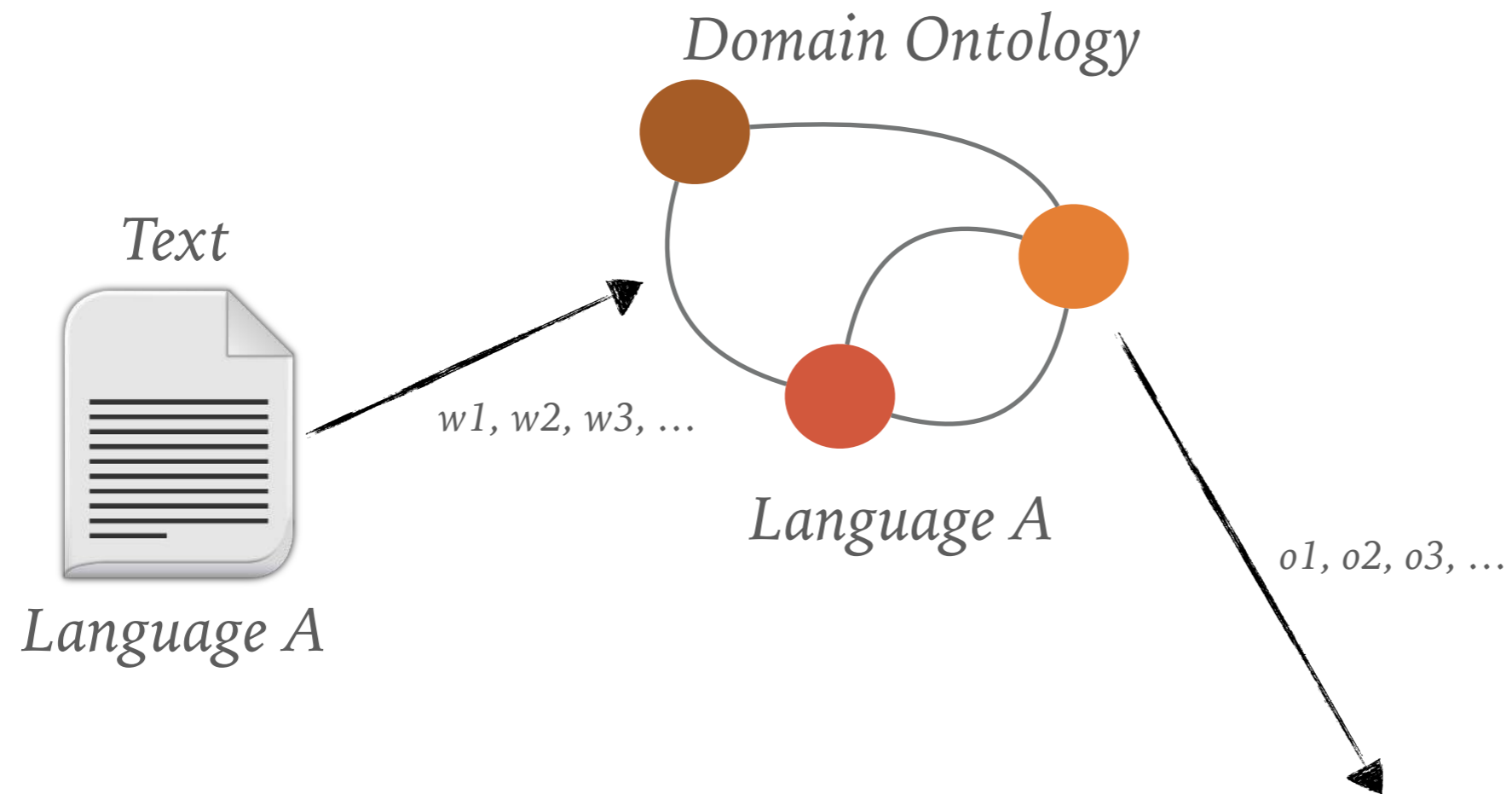


Language A

THE PROBLEM

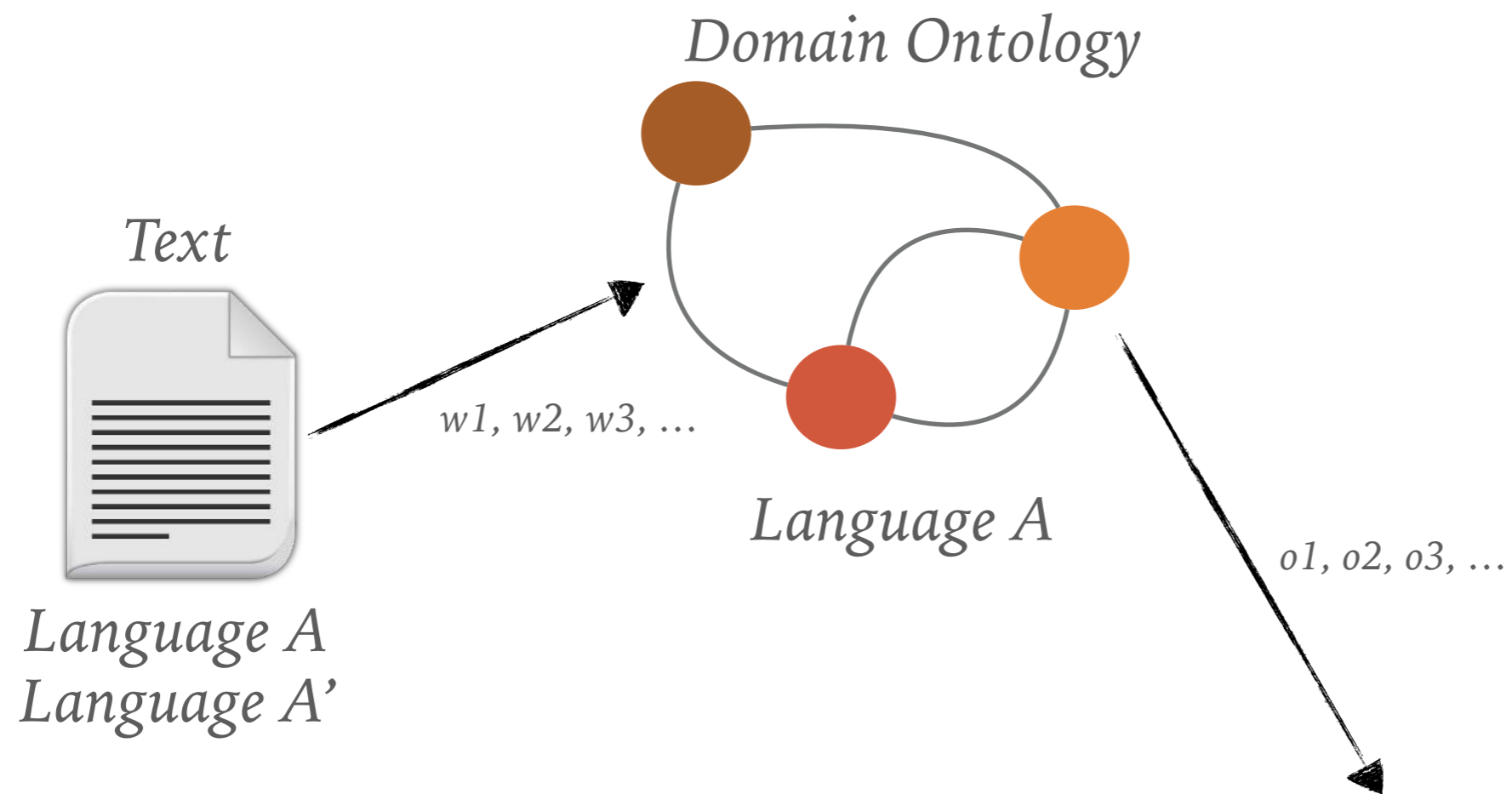


THE PROBLEM



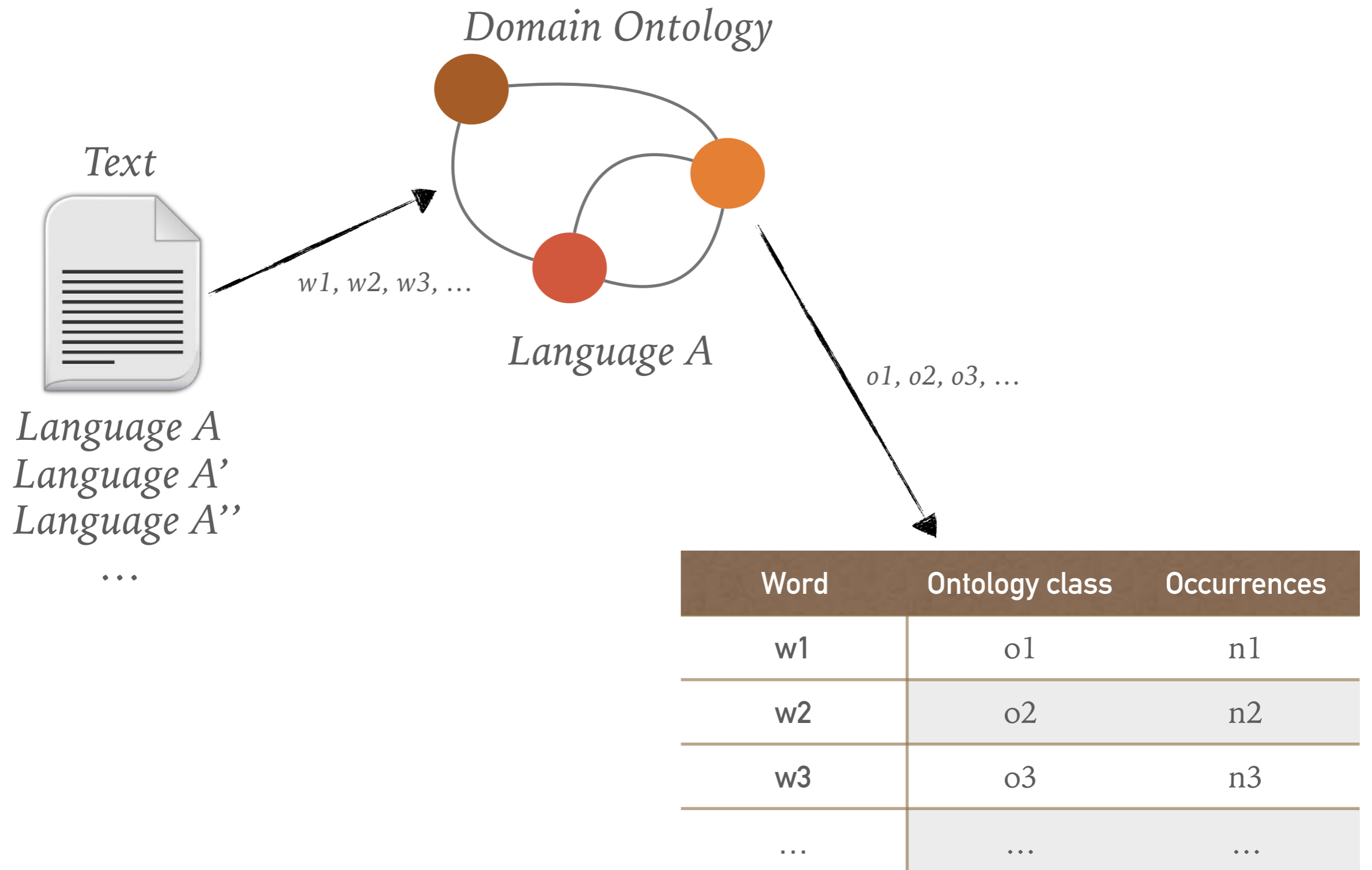
Word	Ontology class	Occurrences
w1	o1	n1
w2	o2	n2
w3	o3	n3
...

THE PROBLEM

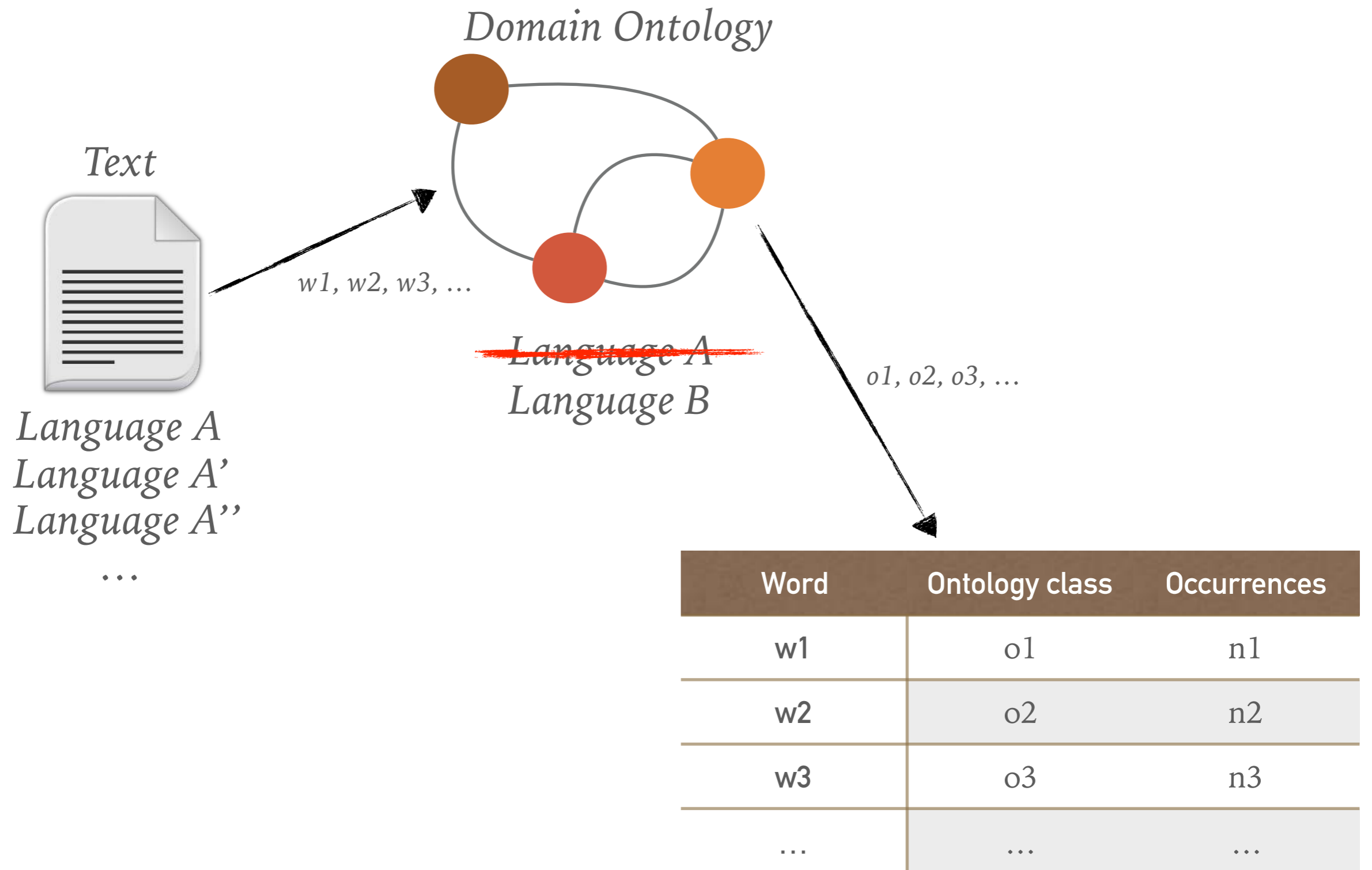


Word	Ontology class	Occurrences
w1	o1	n1
w2	o2	n2
w3	o3	n3
...

THE PROBLEM



THE PROBLEM



HOW TO USE BABELNET TO SOLVE THIS PROBLEM



HOW TO USE BABELNET TO SOLVE THIS PROBLEM

w1

HOW TO USE BABELNET TO SOLVE THIS PROBLEM

$w_1 \longrightarrow (\text{lemma}_1, \text{POS}_1)$

HOW TO USE BABELNET TO SOLVE THIS PROBLEM

$w1 \longrightarrow (lemma1, POS1)$

\downarrow `getSynsets(lemma1, Language A, POS1)`

$\{synset1, synset2, \dots\}$ in *Language A*

HOW TO USE BABELNET TO SOLVE THIS PROBLEM

$w1 \longrightarrow (lemma1, POS1)$

$getSynsets(lemma1, Language A, POS1)$

$\{synset1, synset2, \dots\}$ in **Language A**

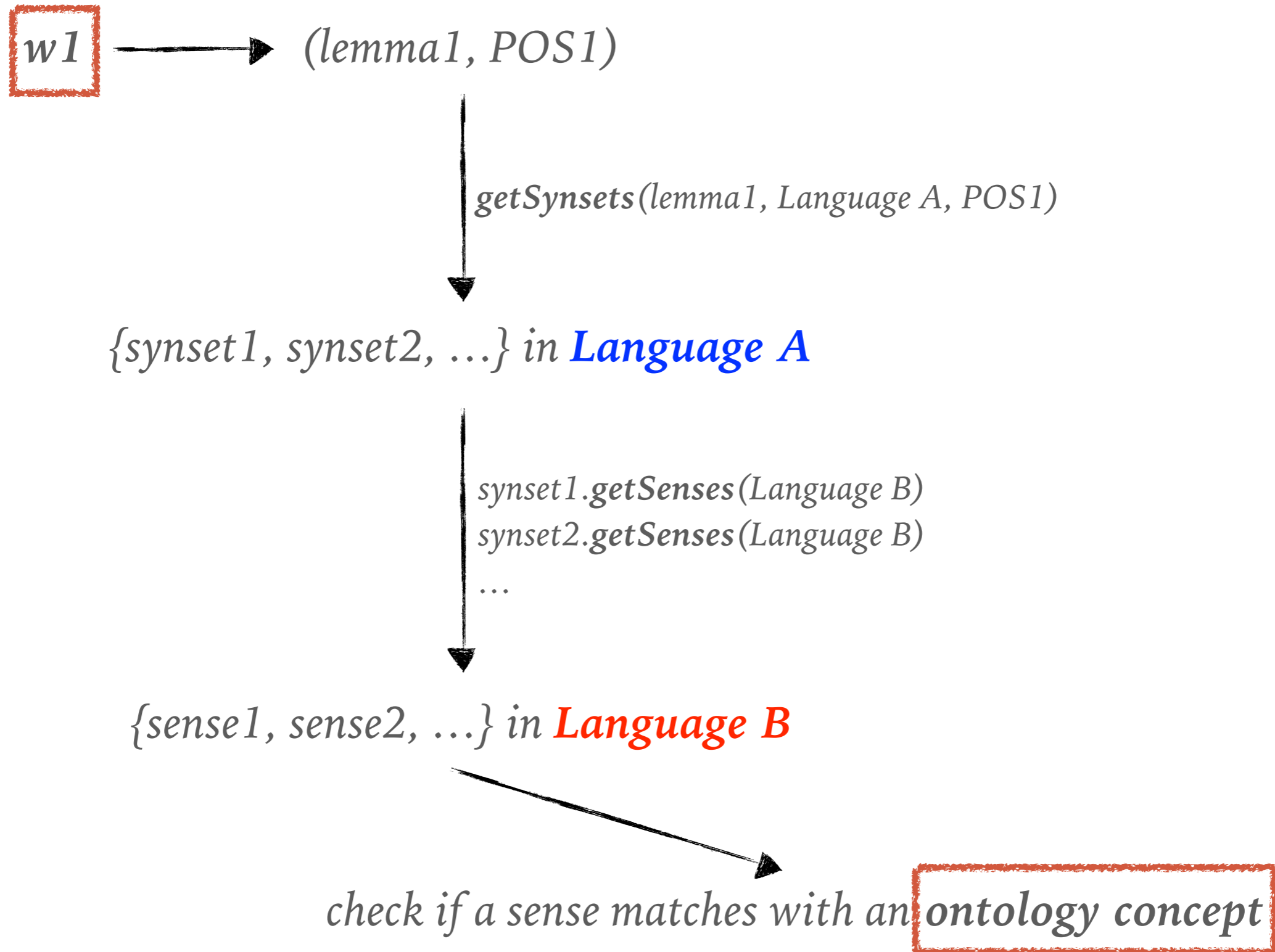
$synset1.getSenses(Language B)$

$synset2.getSenses(Language B)$

...

$\{sense1, sense2, \dots\}$ in **Language B**

HOW TO USE BABELNET TO SOLVE THIS PROBLEM

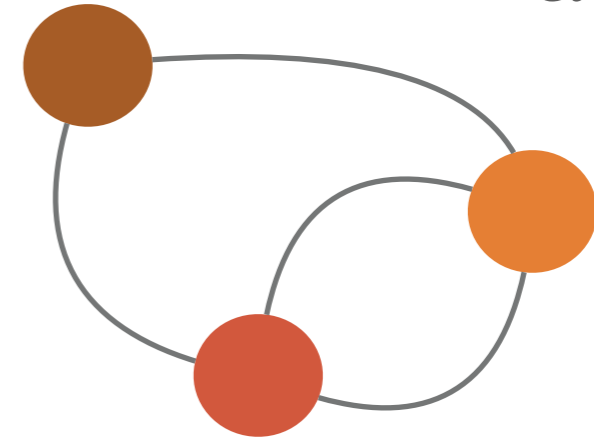


BASIC ONTOLOGY-DRIVEN TEXT CLASSIFIER OVERVIEW

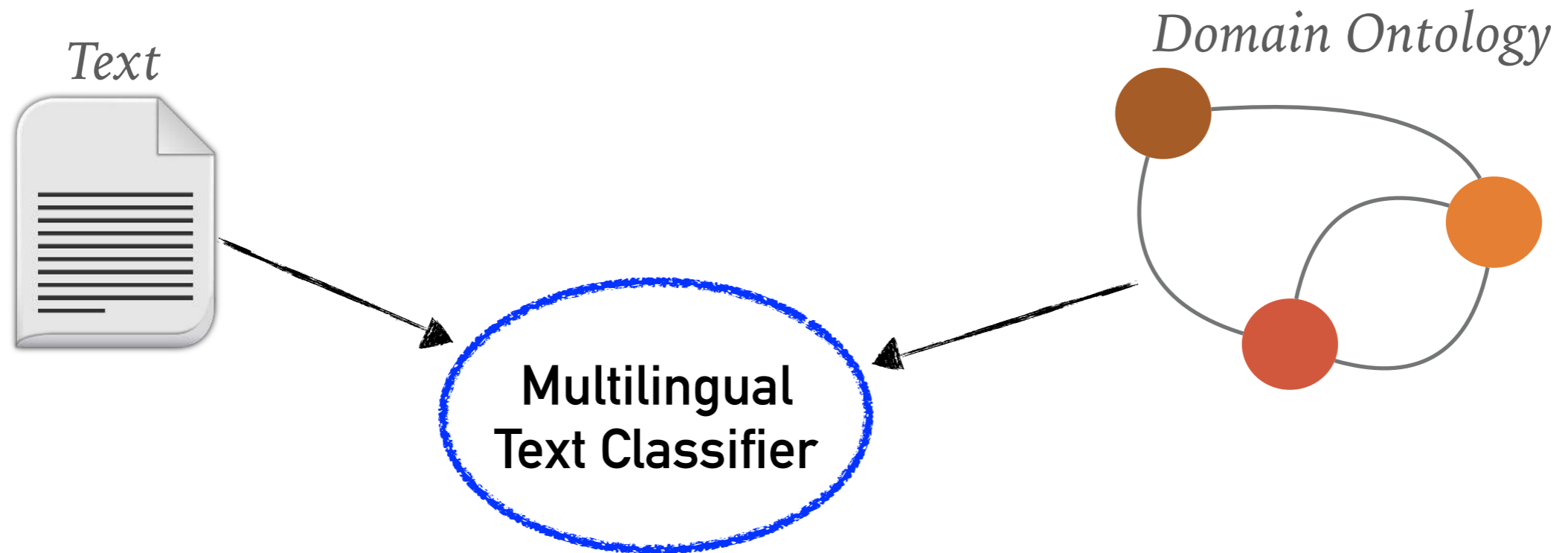
Text



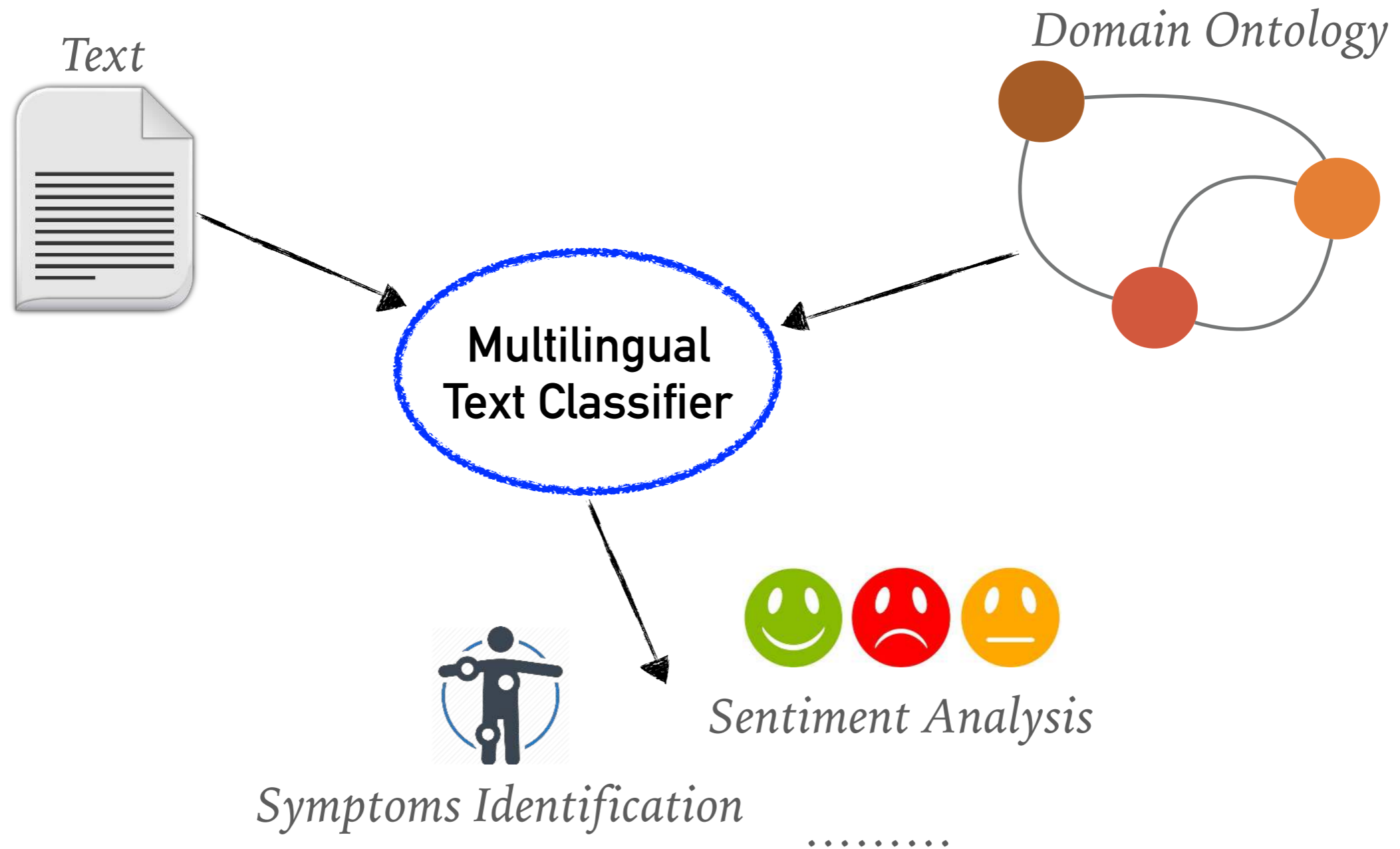
Domain Ontology



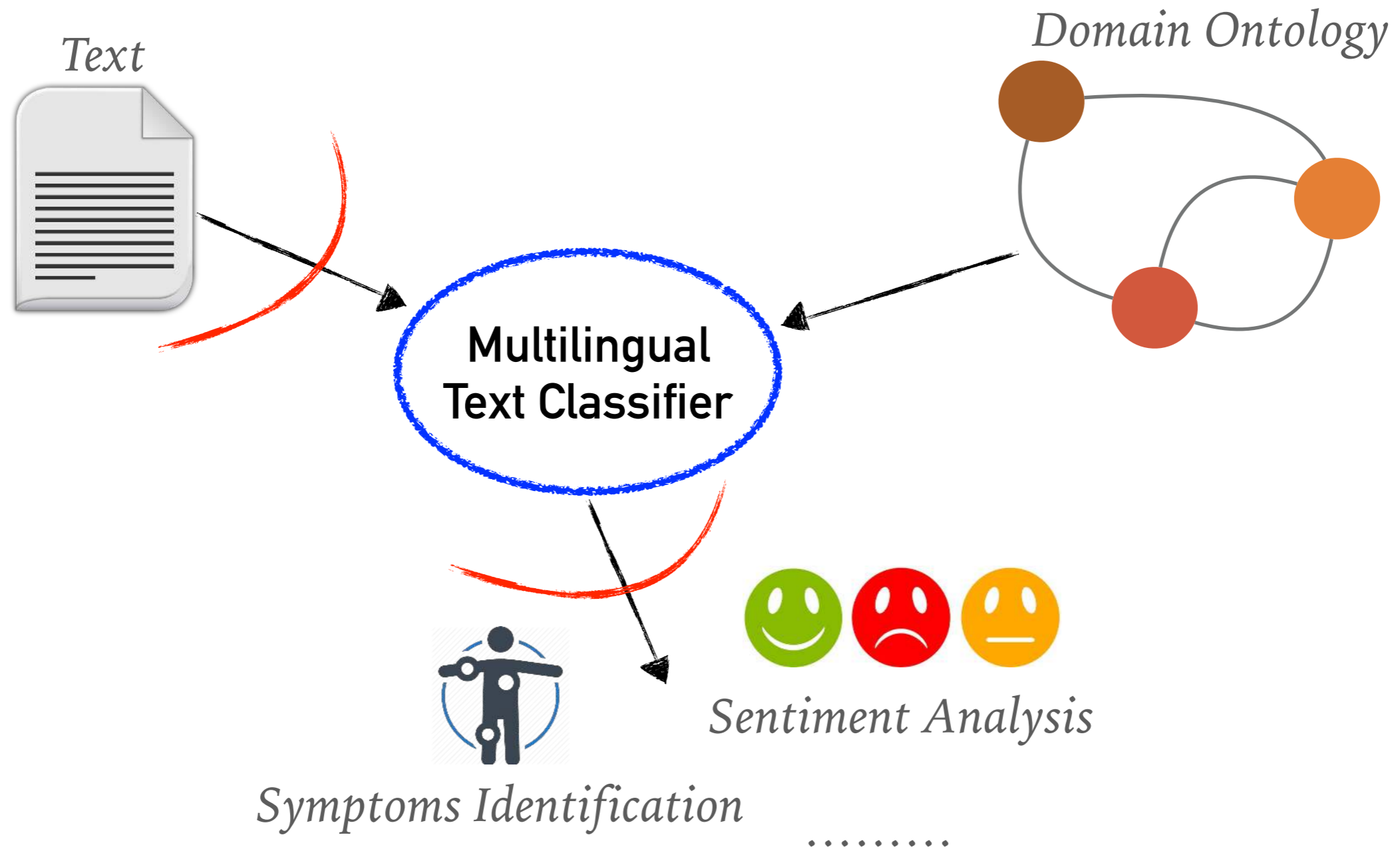
BASIC ONTOLOGY-DRIVEN TEXT CLASSIFIER OVERVIEW



BASIC ONTOLOGY-DRIVEN TEXT CLASSIFIER OVERVIEW



BASIC ONTOLOGY-DRIVEN TEXT CLASSIFIER OVERVIEW



THE MOOD-TC MODULARITY



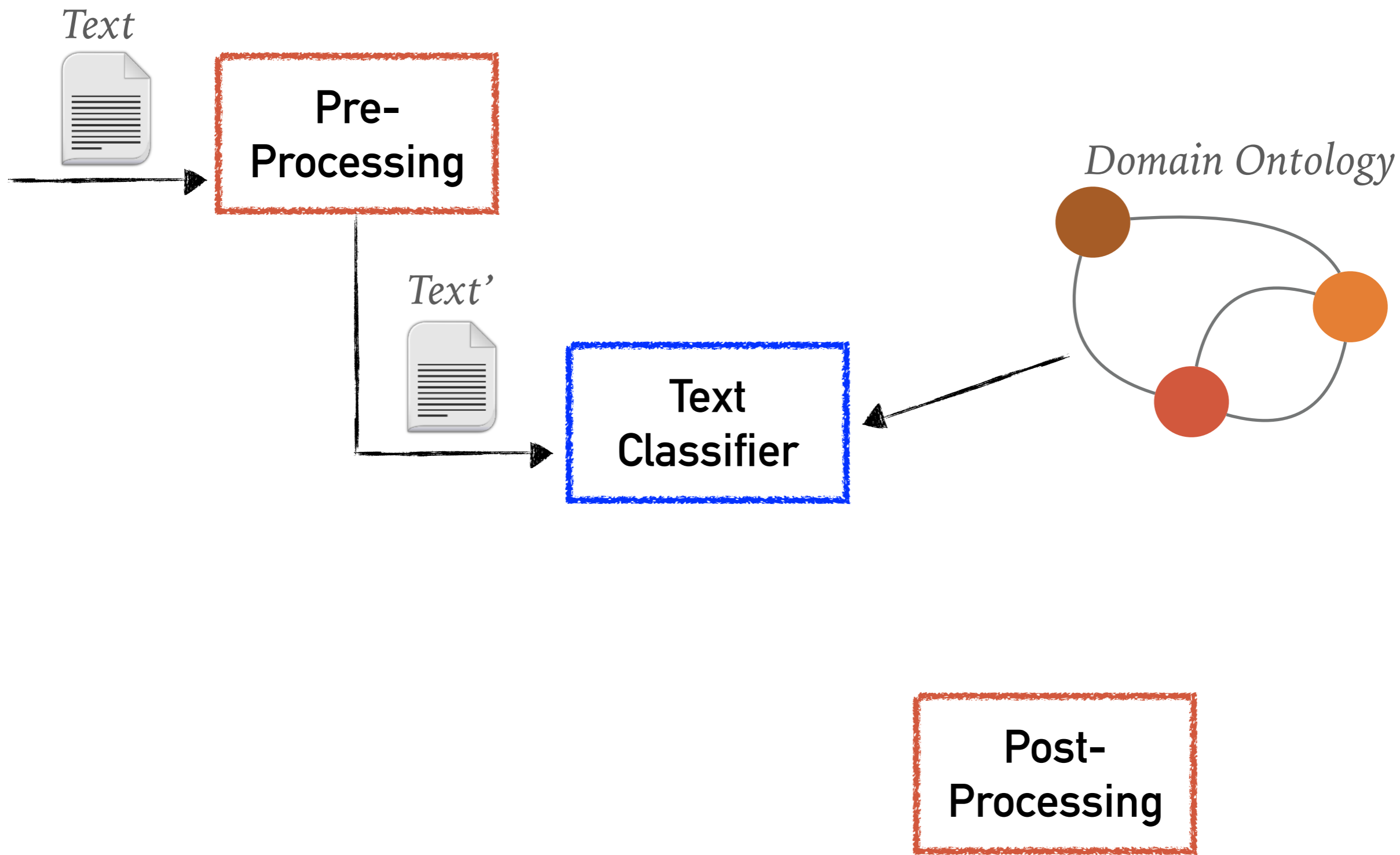
THE MOOD-TC MODULARITY

Pre-
Processing

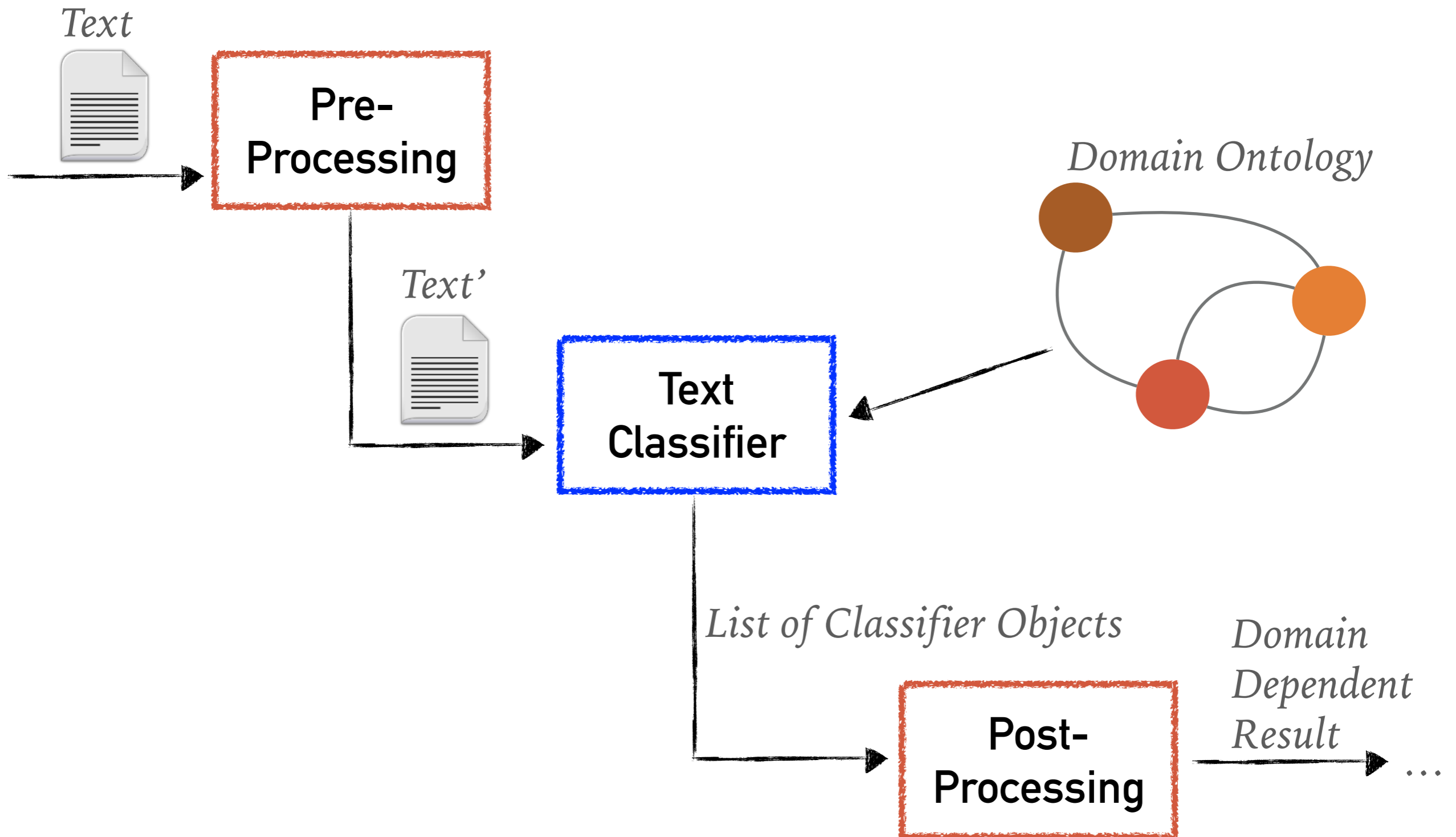
Text
Classifier

Post-
Processing

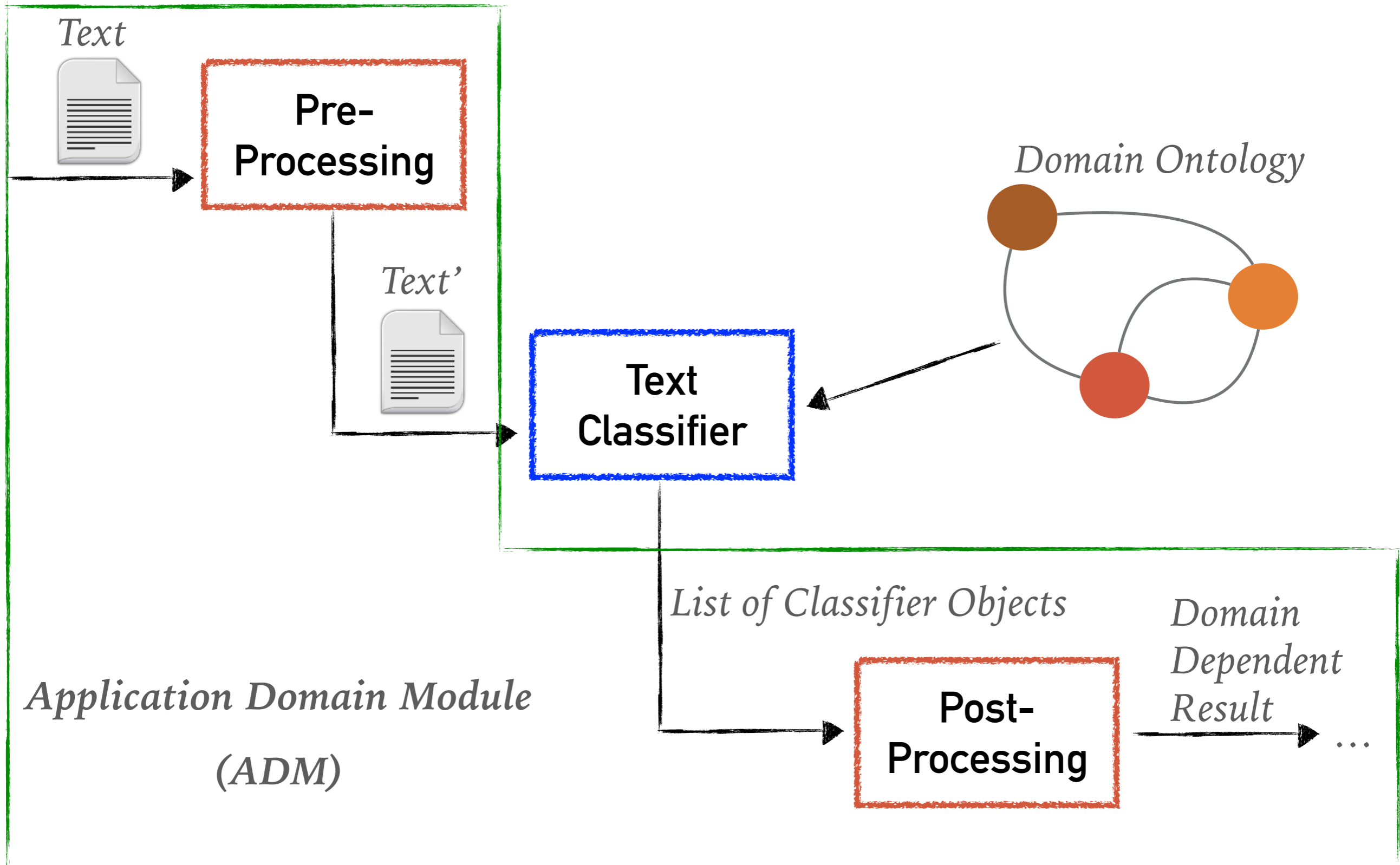
THE MOOD-TC MODULARITY



THE MOOD-TC MODULARITY



THE MOOD-TC MODULARITY



SENTIMENT ANALYSIS RESULTS

- ambiguous-opinion
- negative-opinion
- neutral-opinion
- positive-opinion
- **positive-emotion**
- **positive-accommodation**
- positive-experience-features
- positive-food-features
- positive-location-features
- positive-price_features
- positive-room-features
- positive-staff-features
 - accommodating
 - attentive
 - caring
 - cheerful

MOod-TC
File selection

ONTOLOGY

Language: English

- ⊕ negative-opinion
- ⊕ positive-opinion
 - ⊕ positive-accomm
 - ⊕ positive-food-
 - ⊕ positive-room
 - awesome
 - high-grade
 - spacious
 - high-class
 - fashionable
 - scent
 - beautiful

TEXT

Language: Italian

Gran **bella** esperienza. Certamente da ripetere. Questo hotel è **bellissimo** e di alto livello...decisamente superiore ai diversi 4 stelle visitati in riviera fino ad ora. Una menzione di merito va allo staff di altissimo livello. Dalla ragazza alla reception ai camerieri di sala, senza tralasciare il ragazzo che si occupa del parcheggio auto. Le stanze sono stupende e spaziose. Pulitissime comode. I servizi sono d'eccezione a partire dalla wifi gratuita in tutto l'hotel fino alla apprezzatissima assenza di macchinette per il caffè/cappuccino in sala colazione. **Bella** anche la zona wellness, con la piscina/idromassaggio all'aperto sull'attico, molto **bello**. seppure piccolina è di buon livello e contribuisce a raggiungere il relax che si ricerca in una struttura così. la

Lemma word	Ontology word	Occurences
bello	beautiful	4
superiore	top	1
stupendo	wonderful	1
pulito	clean	1
comodo	comfortable	1
gratuito	free	1
molto	great	1

MODULE

No Module

RESULT

CLASSIFY

Overall Score	Reviews EN			Reviews IT			Reviews FR			Reviews ES			Reviews DE			Reviews		
	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total
	N	%		N	%		N	%		N	%		N	%		N	%	
5	74	91,4	81	5	100,0	5	5	100,0	5	4	80,0	5	5	100,0	5	93	92,1	101
4	145	90,1	161	5	100,0	5	5	100,0	5	4	80,0	5	5	100,0	5	164	90,6	181
3	77	82,8	93	5	100,0	5	4	80,0	5	4	80,0	5	5	100,0	5	95	84,1	113
2	16	31,4	51	3	60,0	5	2	40,0	5	2	40,0	5	0	0,0	5	23	32,4	71
1	39	56,5	69	2	40,0	5	4	80,0	5	4	80,0	5	4	80,0	5	53	59,6	89

SENTIMENT ANALYSIS – EXAMPLE

*Gran **bella** esperienza. Certamente da ripetere.*

*Questo hotel è **bellissimo** e di alto livello...decisamente **superiore** ai diversi 4 stelle visitati in riviera fino ad ora. Una menzione di merito va allo staff di altissimo livello. Dalla ragazza alla reception ai camerieri di sala, senza tralasciare il ragazzo che si occupa del parcheggio auto.*

*Le stanze sono **stupende** e spaziose. **Pulitissime** comode. I servizi sono d'eccezione a partire dalla **wifi gratuita** in tutto l'hotel*

SENTIMENT ANALYSIS - EXAMPLE

Gran **bella** esperienza. Certamente da ripetere.

beautiful

Questo hotel è **bellissimo** e di alto livello...decisamente **superiore** ai diversi 4 stelle visitati in riviera fino ad ora. Una menzione di merito va allo staff di altissimo livello. Dalla ragazza alla reception ai camerieri di sala, senza tralasciare il ragazzo che si occupa del parcheggio auto.

top

Le stanze sono **stupende** e spaziose. **Pulitissime** comode. I servizi sono d'eccezione a partire dalla **wifi gratuita** in tutto l'hotel

wonderful

clean

comfortable

free_wifi

CONCLUSIONS

- The TC gave us good results with very little effort (thanks to BabelNet)
- The multilingual aspect supports the reuse of existent domain ontologies without the necessity to duplicate all concepts for each language
- Our approach's modularity allows reusing the TC in a wide range of problems (s.t Sentiment analysis, Symptom extraction, and so on) simply changing the domain ontology and the pre-/post- processing phase

CONCLUSIONS

- The TC gave us good results with very little effort (thanks to BabelNet)
- The multilingual aspect supports the reuse of existent domain ontologies without the necessity to duplicate all concepts for each language
- Our approach's modularity allows reusing the TC in a wide range of problems (s.t Sentiment analysis, Symptom extraction, and so on) simply changing the domain ontology and the pre-/post- processing phase

FUTURE WORKS

- To improve the POS processing in order to have more detailed and useful information which allow optimising the search within BabelNet
- To manage the negation, the irony and other important POS aspects which have to be resolved in order to query BabelNet correctly
- To add Word Sense Disambiguation in order to reduce the number of BabelSynsets returned by BabelNet