Multilingual Sentiment Analysis

Viviane Moreira

viviane@inf.ufrgs.br

UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

.inf
INSTITUTO
DE INFORMÁTICA
UFRGS

Context

- Sentiment Analysis/Opinion Mining is the field of study that **analyses people's opinions** expressed in written language (frequently in the form of **reviews**)

- One of the most common types of analysis consists in determining the **polarity** of a review (i.e. whether it is positive or negative)

- Different levels of granularity (review, sentence, **aspect**)

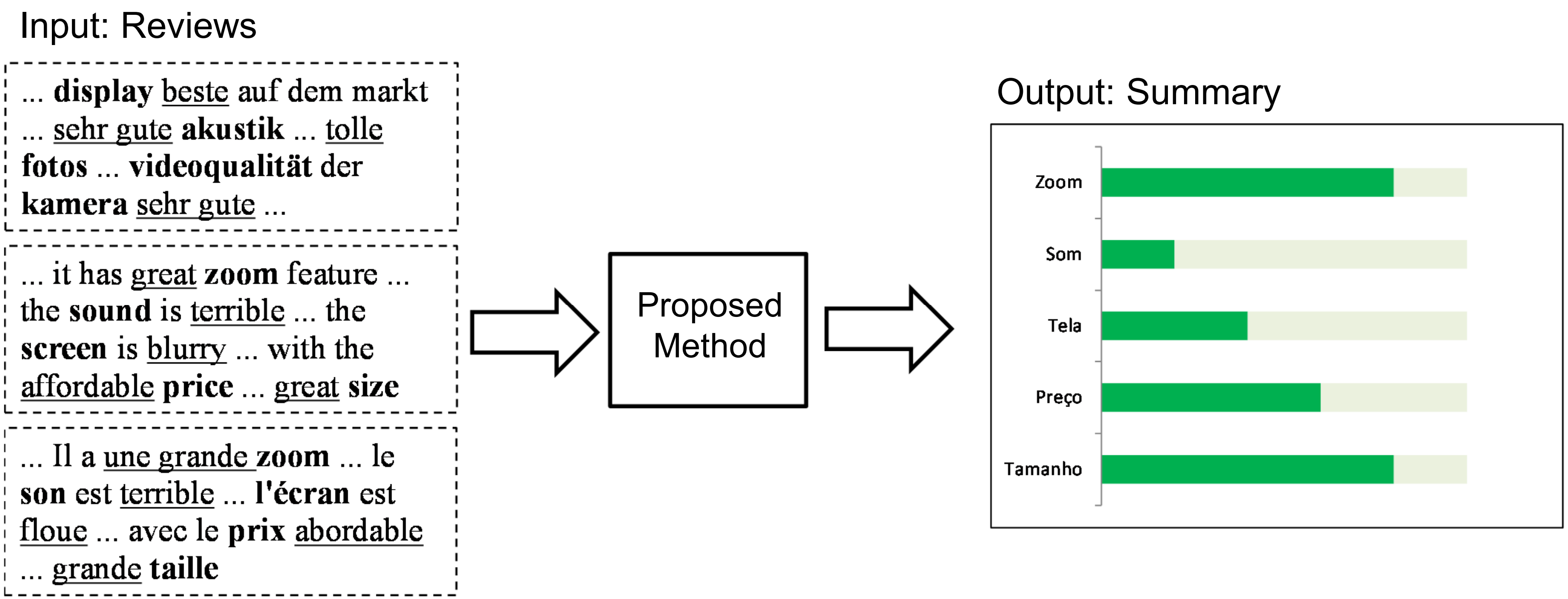
Motivation and Goal

- The number of reviews in amazon.com is, on average, 7 times higher than the sum of the reviews in the 6 major Brazilian e-commerce web sites.
- Address the lack of product reviews in Brazilian web sites

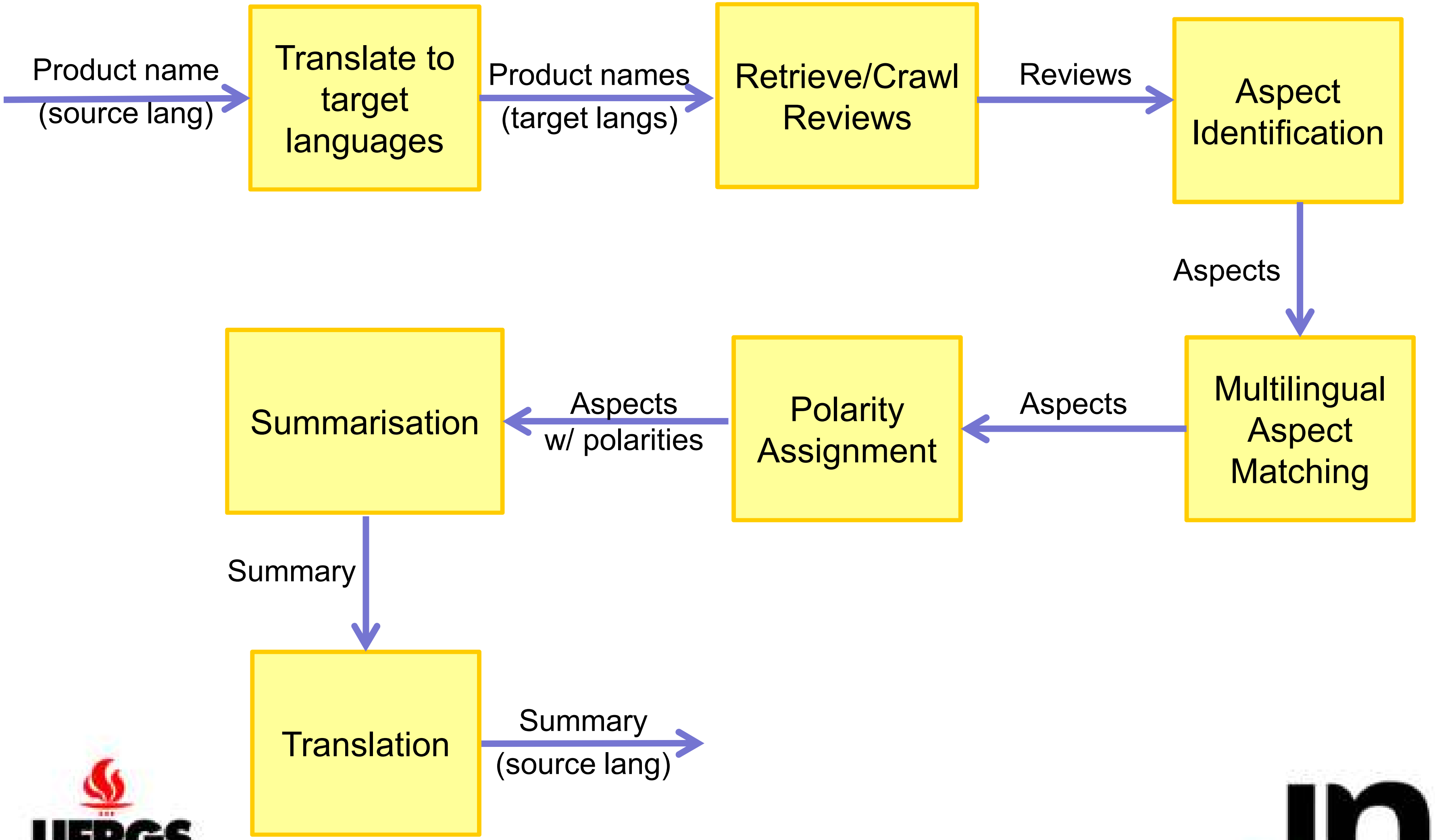
UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

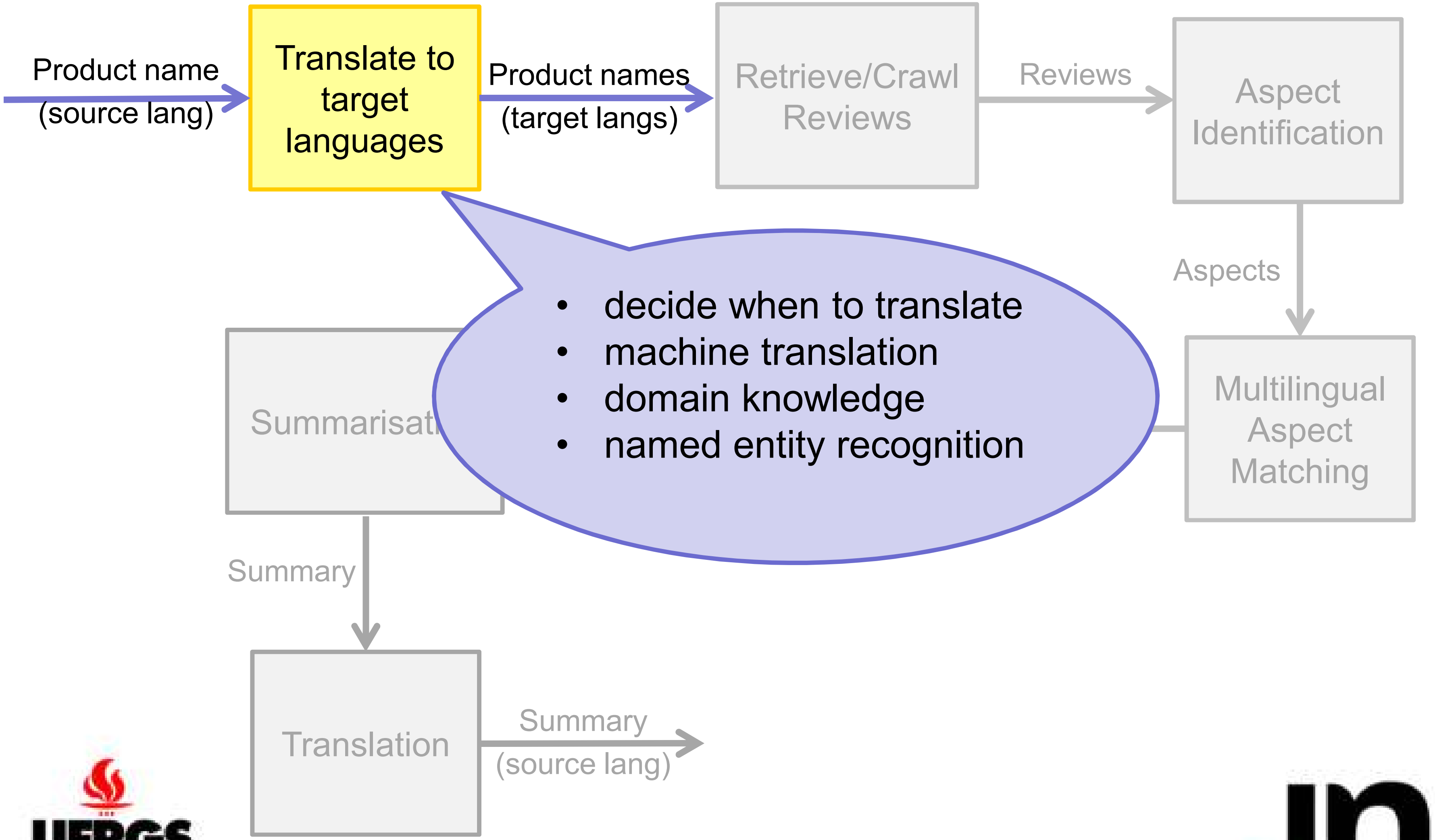
Goal


Input: Reviews
... display beste auf dem markt ... sehr gute akustik ... tolle fotos ... videoqualität der kamera sehr gute ...
... it has great zoom feature ... the sound is terrible ... the screen is blurry ... with the affordable price ... great size
... Il a une grande zoom ... le son est terrible ... l'écran est floue ... avec le prix abordable ... grande taille
Proposed Method
Output: Summary
Zoom
Som
Tela
Preço
Tamanho

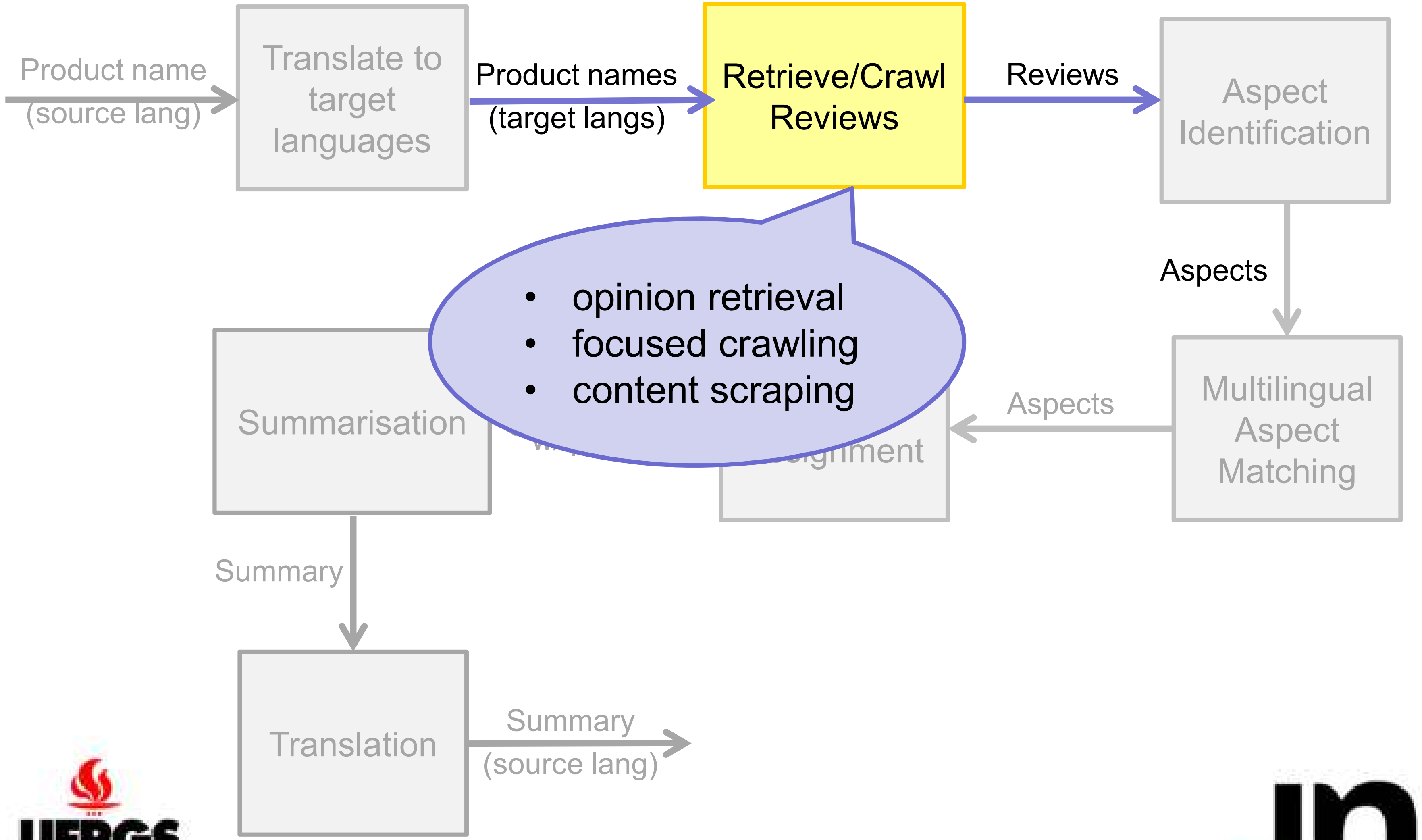
Overview


Product name
(source lang)
Translate to target languages
Product names
(target langs)
Retrieve/Crawl Reviews
Reviews
Aspect Identification
Aspects
Multilingual Aspect Matching
Aspects
Polarity Assignment
Aspects
w/ polarities
Summarisation
Summary
Translation
Summary
(source lang)

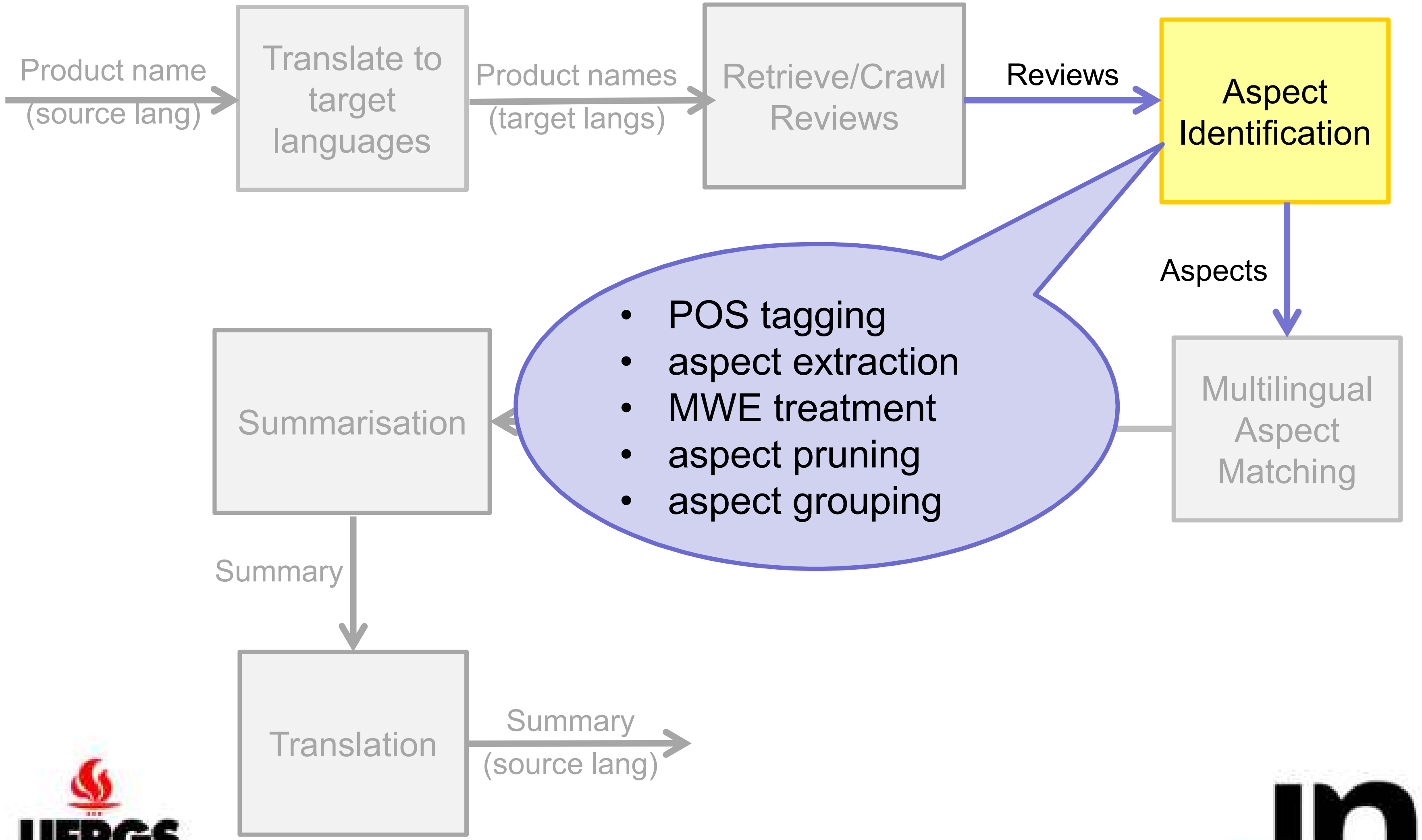
Activities/resources involved in each step



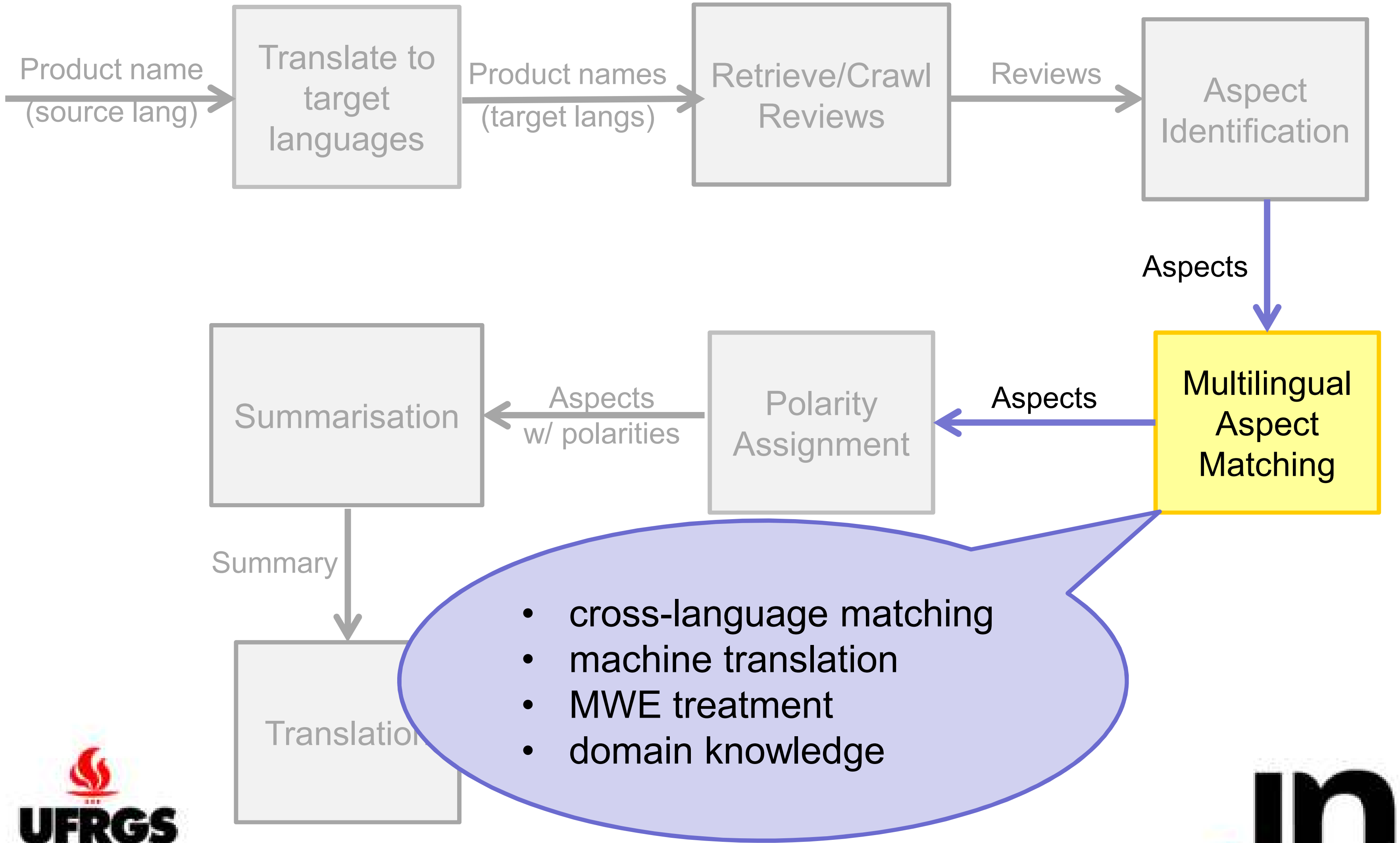
Activities/resources involved in each step



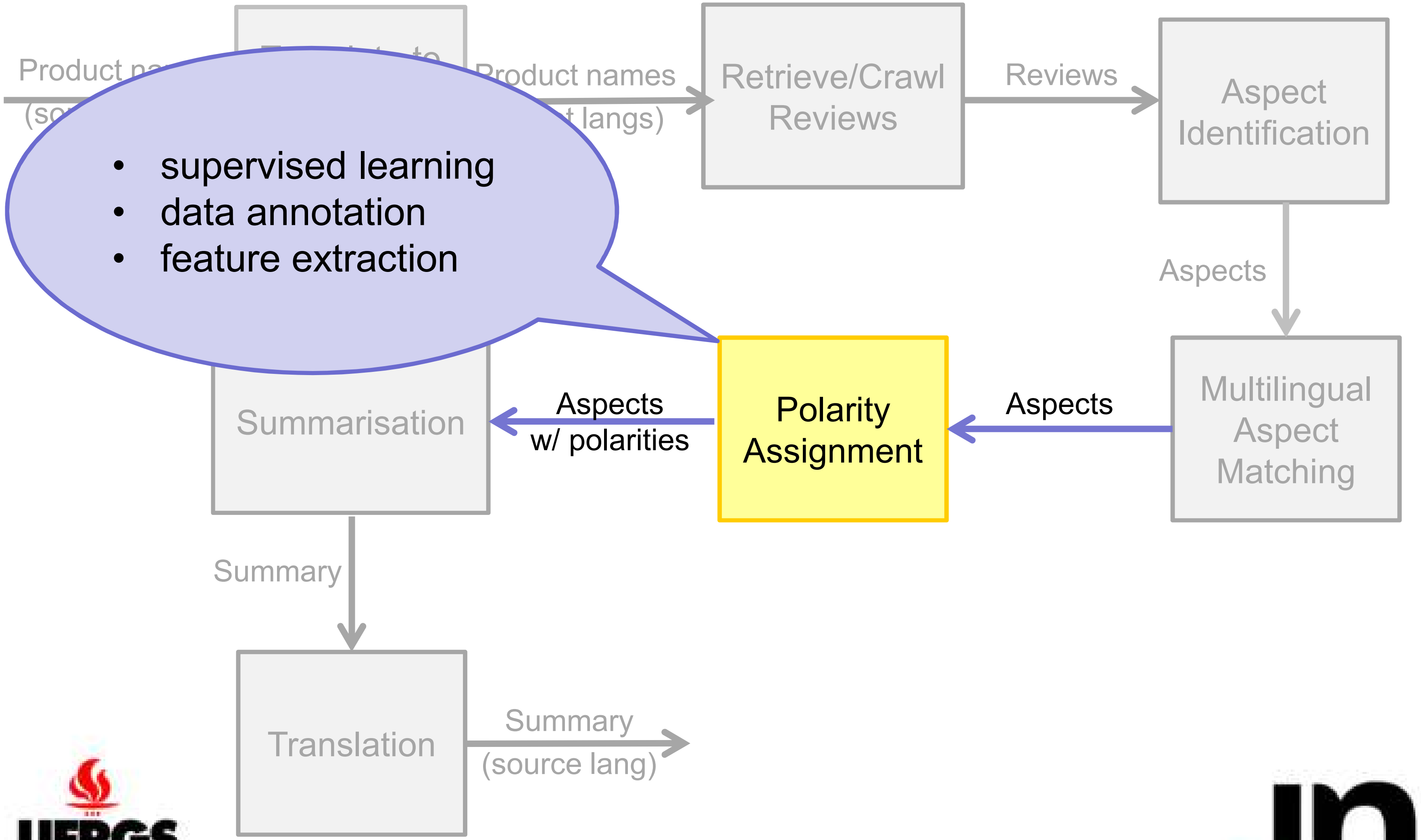
Activities/resources involved in each step



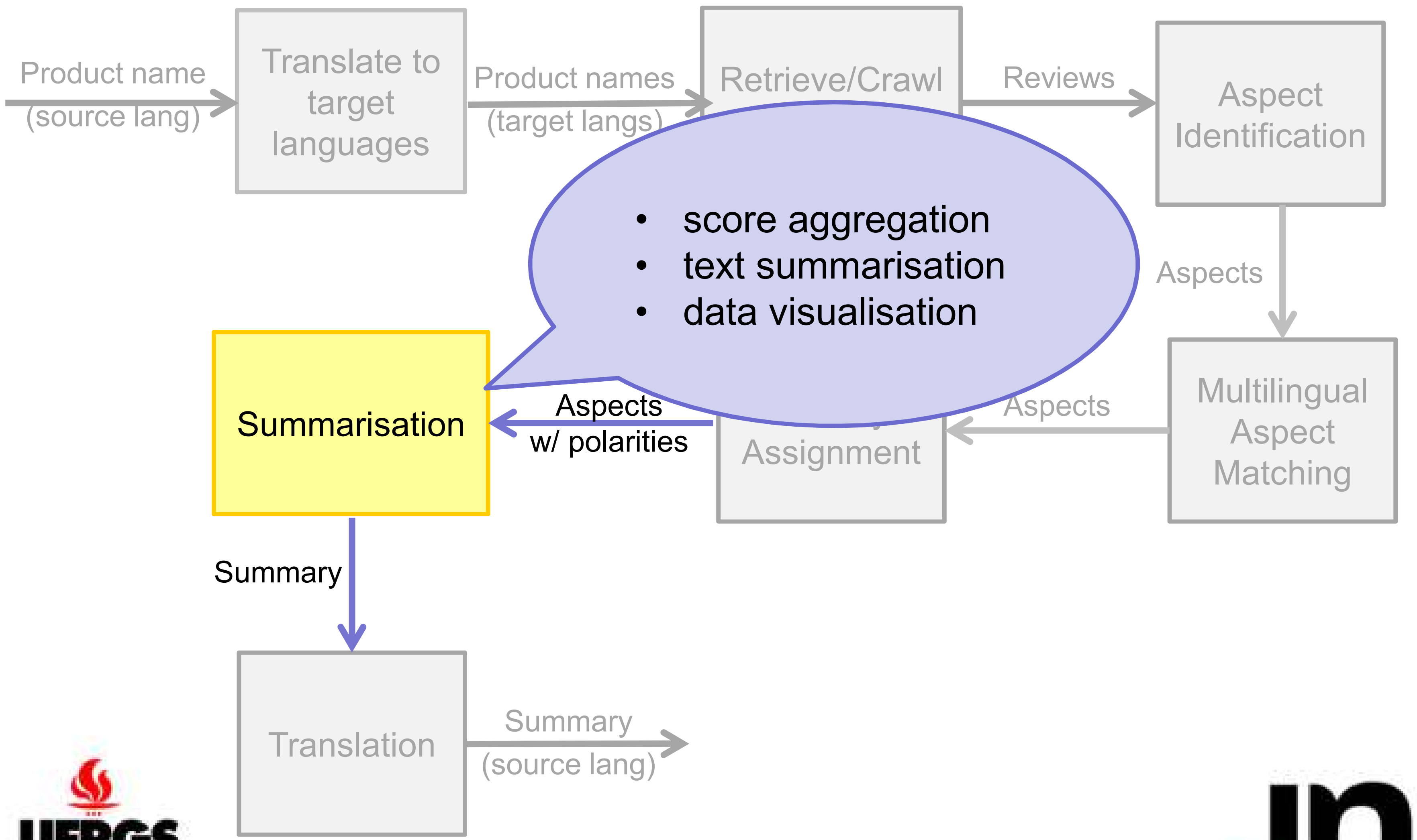
Activities/resources involved in each step



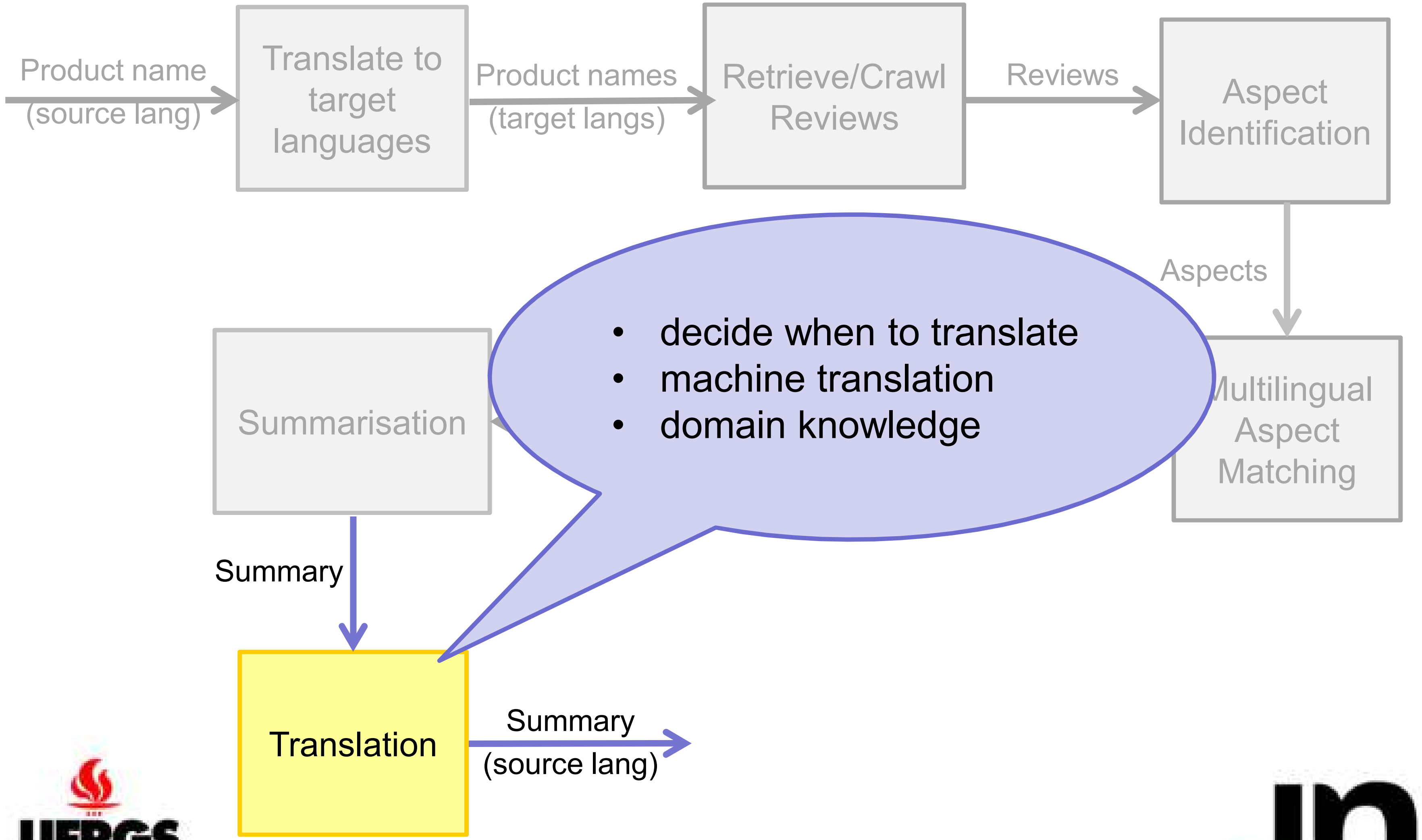
Activities/resources involved in each step



Activities/resources involved in each step



Activities/resources involved in each step



Previous Work Related to this Project

- Polarity classification relying on features derived from Information Retrieval
  - MSc by Anderson Kauer (concluded in 2016)
  - Publication: Anderson Uilian Kauer, Viviane Pereira Moreira. Using information retrieval for sentiment polarity prediction. Expert Syst. Appl. 61: 282-289 (2016).

- Multilingual Schema Matching
  - Publication: Thanh Hoang Nguyen, Viviane Pereira Moreira, Huong Nguyen, Hoa Nguyen, Juliana Freire: Multilingual Schema Matching for Wikipedia Infoboxes. PVLDB 5(2): 133-144 (2011)

Using Information Retrieval for Sentiment Polarity Prediction

SABIR - Sentiment Analysis Based on Information Retrieval

MSc by Anderson Kauer

(concluded in 2016)

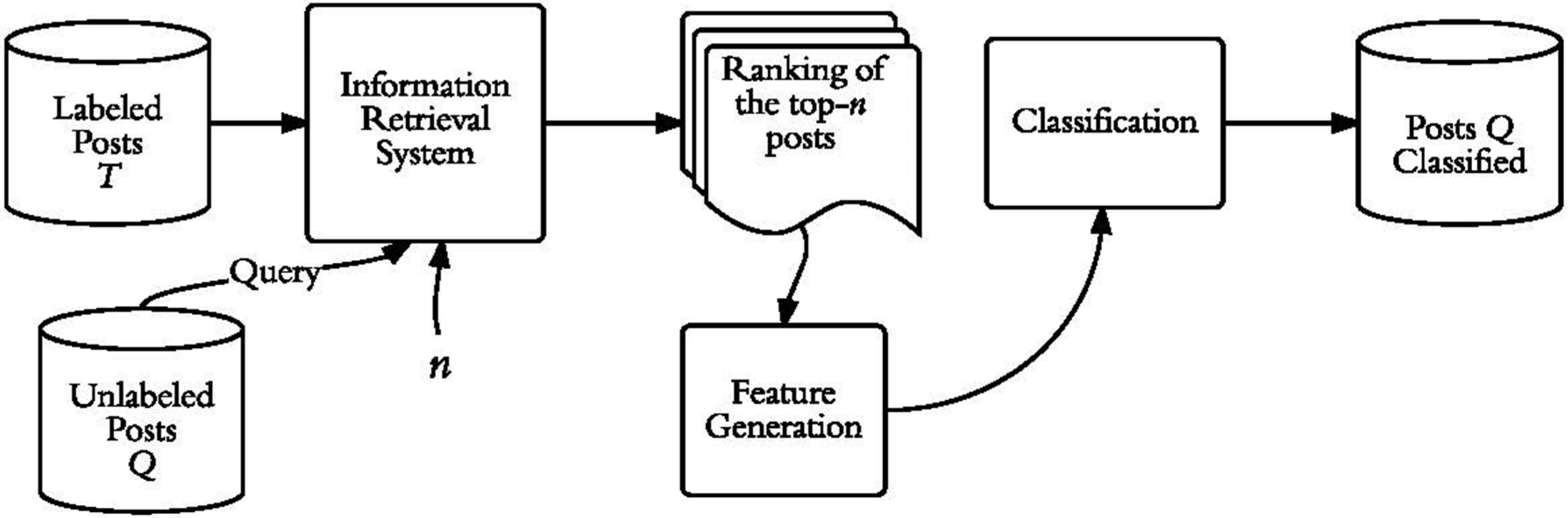
UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Overview

- Application scenario: Classify the polarity of tweets
- Most of the existing approaches use a **BoW** classifier: **sparse** features
- Our features are derived from the ranking generated by an Information Retrieval System in response to a query *q* which consists of the tweet that we wish to classify

Overview



More formally:

Given a set of tweets $T = \{t_1, t_2, \ldots, t_m\}$ for which the class $c_i \in \{+,-\}$ is known and a set of unlabelled tweets $Q = \{q_1, q_2, \ldots, q_p\}$, we use information about the similarity of each element $q_i \in Q$ in relation to the elements $t_j \in T$ to predict the class of q_i.

UFRGS UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Information Retrieval-based features

Rank $_{abs}$	Score	Class
1	25,897	negative
2	25,541	positive
3	25,158	negative
4	22,729	negative
5	22,204	negative
6	21,576	negative
7	21,468	negative
8	21,426	positive
9	21,011	positive
10	20,985	positive
11	20,666	negative
12	20,666	negative
13	20,546	positive
14	20,546	positive
15	20,435	negative
16	20,07	negative
17	20,032	negative
18	20,025	positive
19	19,974	positive
20	19,962	positive
21	19,934	positive
22	19,902	positive
23	19,889	positive
24	19,779	negative
25	19,771	positive

Positive

Rank $_{abs}$	Rank $_{rel}$	Score
2	1	25,541
8	2	21,426
9	3	21,011
10	4	20,985
13	5	20,546
14	6	20,546
18	7	20,025
19	8	19,974
20	9	19,962
21	10	19,934
22	11	19,902
23	12	19,889
25	13	19,771

Feature	Value
avg	20,732
max	25,541
min	19,771
sum	269,51
count	13
ω	5,5743
positional$_\omega$	115,51
avg$_\omega$	0,2688
max$_\omega$	0,2182
min$_\omega$	0,2819
sum$_\omega$	0,0206
count$_\omega$	0,4287

Negative

Rank $_{abs}$	Rank $_{rel}$	Score
1	1	25,897
3	2	25,158
4	3	22,729
5	4	22,204
6	5	21,576
7	6	21,468
11	7	20,666
12	8	20,666
15	9	20,435
16	10	20,07
17	11	20,032
24	12	19,779

Feature	Value
avg	21,723
max	25,897
min	19,779
sum	260,68
count	12
ω	8,5822
positional$_\omega$	188,44
avg$_\omega$	0,395
max$_\omega$	0,3313
min$_\omega$	0,4338
sum$_\omega$	0,0329
count$_\omega$	0,7151

Experiments

- Datasets

Dataset	# tweets	# positive	# negative	# unigrams
STD-train	1.6M	800K	800K	480785
STD-test	359	182	177	1683
HCR-train	614	213	401	2869
HCR-test	658	154	504	2927
OMD	1488	541	947	3411
STS-Gold	2034	632	1402	5139

- IR System: Zettair using BM25 for ranking, *n*=1000
- Classification Algorithm: Maximum Entropy
- Two versions:
  - SABIR (using the same dataset for training – ignoring the tweet to be classified)
  - SABIR-noisy (using STD-train for training – large but noisy)

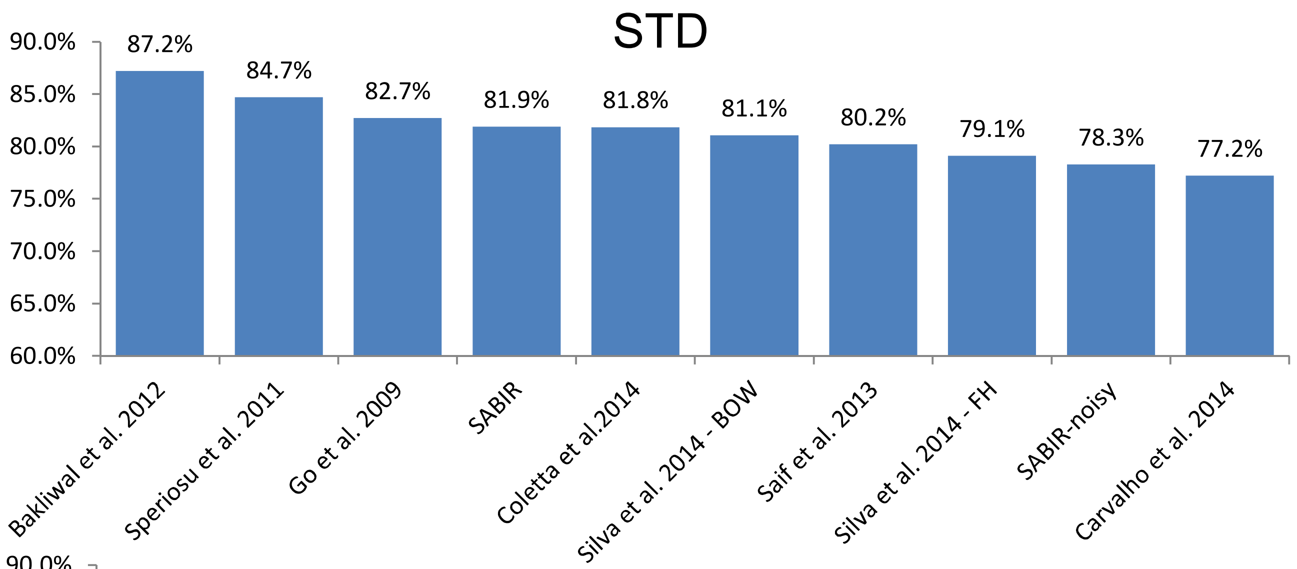
UFRGS UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
.inf UFRGS

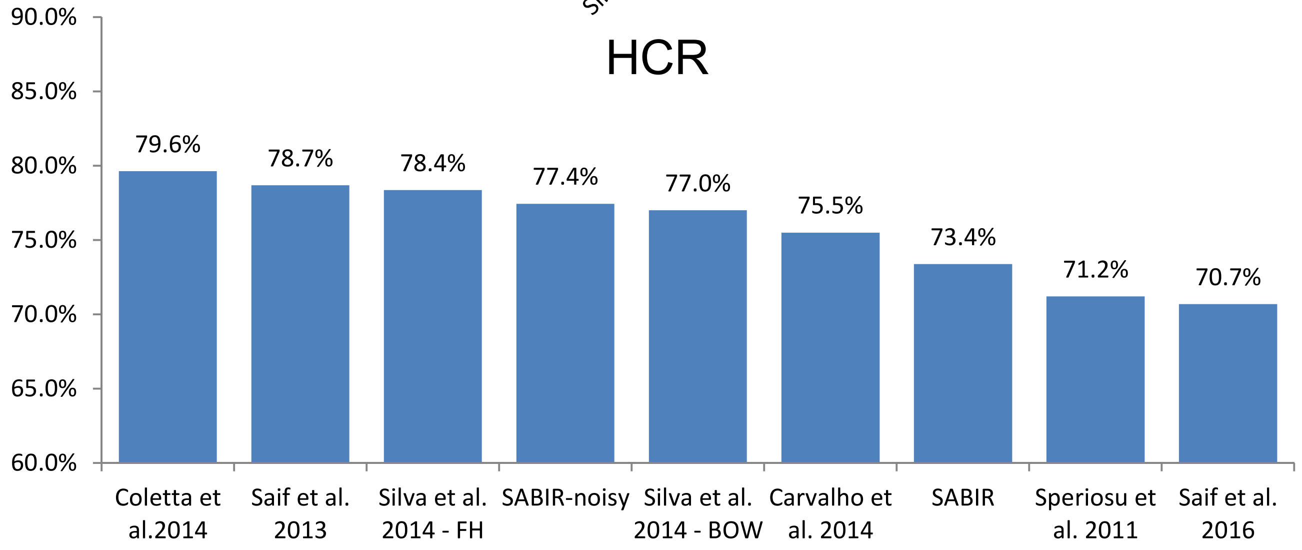
Comparison against BoW baselines

		NB	MNB	SVM	MaxEnt	SABIR-noisy	SABIR
STD	Acc	75.8$^{\diamond\triangledown}$	80.8$^{\diamond\diamond}$	74.7$^{\diamond\triangledown}$	78.6$^{\diamond\diamond}$	78.3$^{\diamond}$	**81.9**
	Rec	75.8	80.8	74.7	78.6	78.3	**81.9**
	Pr	75.8	80.8	74.7	78.6	78.3	**82.0**
	F1	75.7	80.8	74.6	78.6	78.3	**81.9**
HCR	Acc	73.2$^{\diamond\diamond}$	**79.1**$^{\diamond\triangle}$	75.0$^{\diamond\diamond}$	68.0$^{\triangledown\triangledown}$	77.4$^{\diamond}$	73.4
	Rec	73.2	**79.1**	75.8	68.0	77.4	73.4
	Pr	74.7	76.5	**77.1**	76.9	75.7	73.8
	F1	73.9	75.9	75.9	**76.3**	74.6	73.6
OMD	Acc	74.1$^{\triangle\diamond}$	76.5$^{\triangle\diamond}$	77.0$^{\triangle\diamond}$	74.1$^{\triangle\diamond}$	68.4$^{\triangledown}$	**77.5**
	Rec	74.1	76.5	77.0	74.1	68.4	**77.5**
	Pr	73.7	76.2	**77.4**	75.5	67.5	77.1
	F1	73.8	76.3	**77.1**	74.4	64.5	77.0
STS-Gold	Acc	79.7$^{\triangledown\triangledown}$	**84.8**$^{\diamond\diamond}$	84.7$^{\diamond\diamond}$	76.5$^{\triangledown\triangledown}$	83.1$^{\diamond}$	84.5
	Rec	79.7	**84.8**	84.7	76.5	83.1	84.5
	Pr	80.1	**84.6**	**84.6**	78.8	82.8	84.2
	F1	79.9	84.2	**84.6**	77.2	82.9	84.3

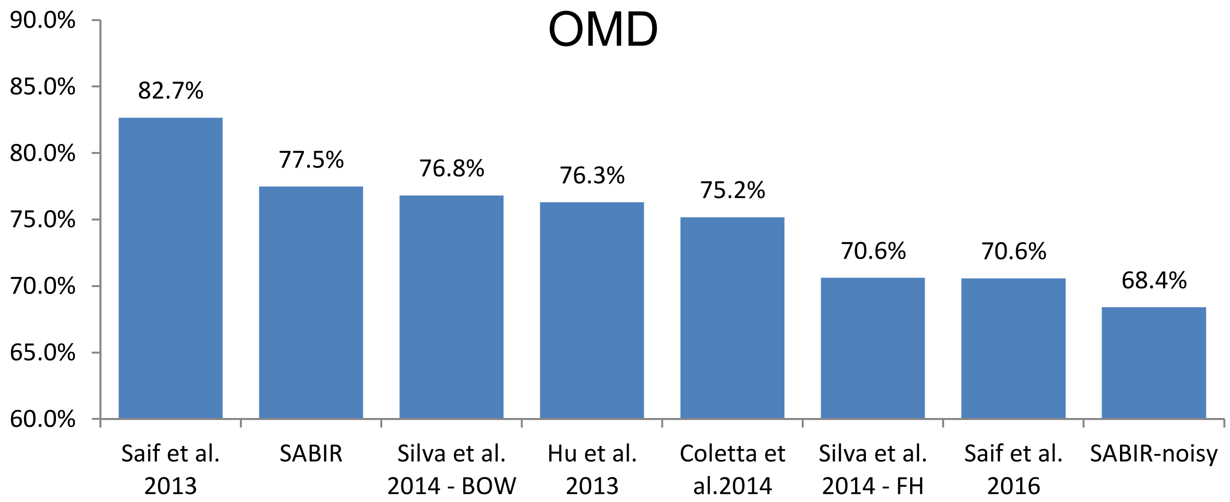
◊ No difference   △ Better than Sabir   ▽ Worse than Sabir

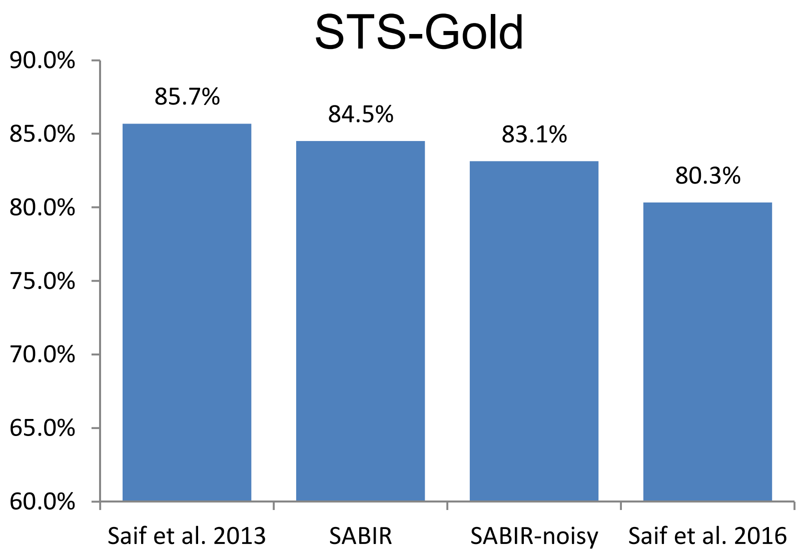
Comparison against published results





Comparison against published results





UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Summary

- Our classifiers built using only 24 features outperformed BoW baselines and are comparable to the state of the art approaches
- Error analysis:
  - posts with interrogations that have sentiment words but in which these words do not express a sentiment;
  - presence of irony; and
  - posts in which there is a negation that inverts the polarity of the sentiment.
- Future work:
  - Other ranking functions
  - Filtering misclassified instances
  - Apply to other classification tasks

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Visualization of review data

Ongoing work in collabotation
with Joao Comba

MSc by Fabian Colque
(to be concluded in 2017)

Demo

Link to the video

Other ongoing research topics (not related to this project)

- Hate speech detection
  - Identify hate speech on the Web
  - Focus on news comments
- Contradiction detection
  - Focused on sentiment-based contradictions
- Crawling/Extraction of CS Conference dates
  - Populate an online database and allow queries

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Ideas for collaboration?

Thank You!