Grégoire JADI

gregoire.jadi@univ-nantes.fr

Spécialité : Informatique
Laboratoire : LINA
Équipe : TALN
Directeur de thèse : Béatrice Daille
Co-encadrants : Laura Monceaux,
Vincent Claveau

Evaluating Lexical Similarity to Build Sentiment Analysis

Annotated Data

Resources

- Lexicons
- Clusters
- Distributionnal thesaurus
- Vector Space Models

MACHINE LEARNING MODEL

(SVM, Neural Network...)

The Context

Ongoing work

Use analogy solving capabilities of word representations to adresse the issue encountered with distance measure

The results

Positive but weak correlation between reference list and generated list

Works better on some dimension than others

Lexicons do not fully agree between themselves

"The food was awesome but the server was a bit grumpy"

Restaurant : positive
Food : positive
Server : negative

The question
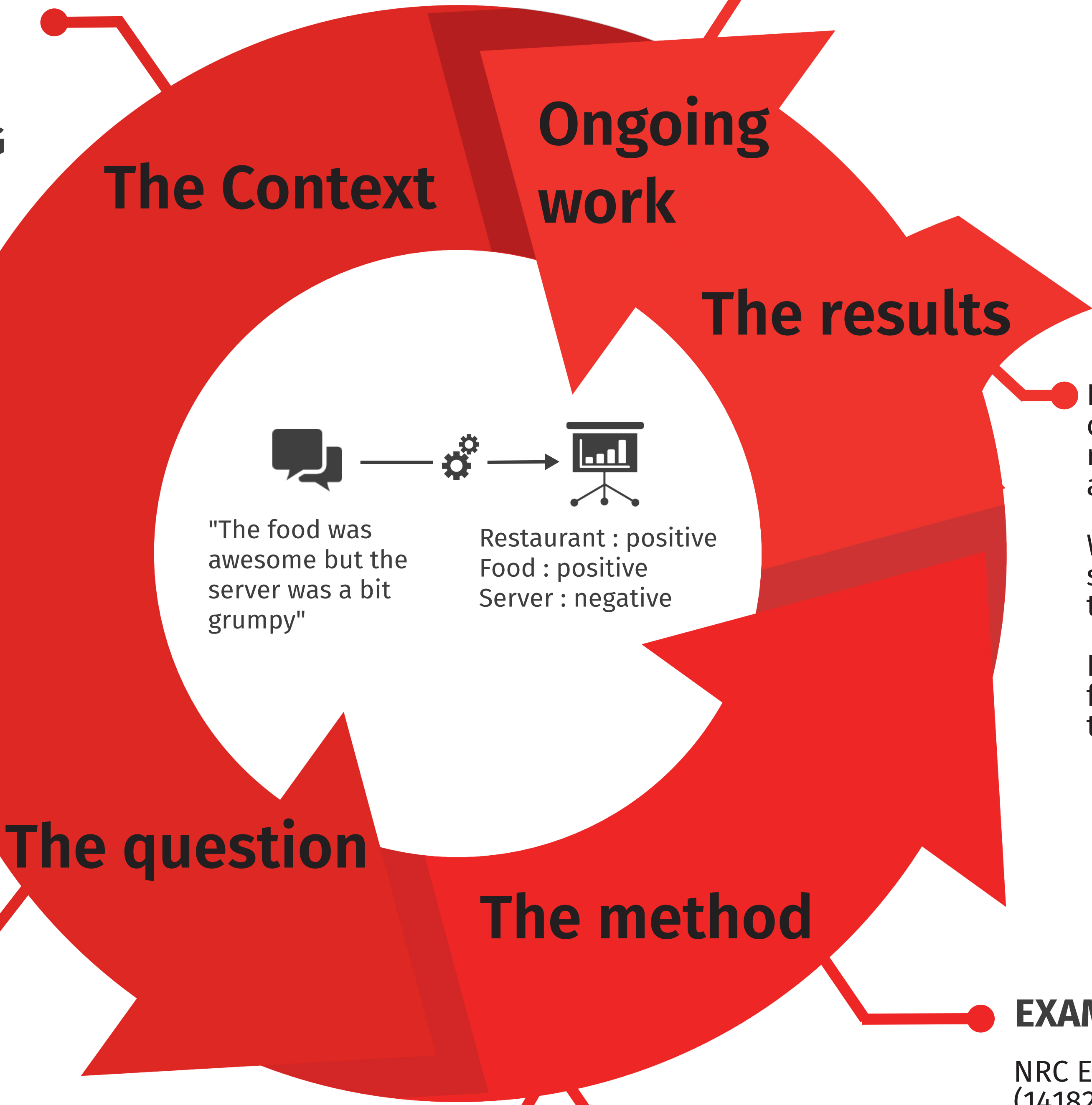
HOW TO EVALUATE RESSOURCES ?

- The classical way :
Train a machine learning model and evaluate its performance

- Our way :
Evaluate resources against manually built lexicons

The method

EXAMPLE

NRC Emotion Lexicon
(14182 words)

PROCESS

1. Generate a reference list from reference lexicons

2. Try to reproduce the list with the word representations

3. Compare the predicted list with the reference list

LEXICONS USED

- ANEW :
3 real valued dimensions
- SentiWordNet :
3 relad valued dimensions
- NRC Emotion Lexicon :
8 binary dimensions

	Abandon	Happy	Weekly
Anger	✗	✗	✗
Anticipation	✗	✓	✗
Disgust	✗	✗	✗
Fear	✓	✗	✗
Joy	✗	✓	✗
Sadness	✓	✗	✗
Surprise	✗	✗	✗
Trust	✗	✓	✗