Towards Cloud-Based Knowledge Capturing Based on Natural Language Processing

Realizing Applied Gaming Ecosystem

Christian Nawroth (University of Hagen)
Matthäus Schmedding (University of Hagen)
Holger Brocks (Research Institute for Communication and Cooperation, Dortmund)
Michael Kaufmann (Lucerne University of Applied Sciences and Arts)
Michael Fuchs (see below)
Matthias Hemmje (University of Hagen)

Michael Fuchs

Professor and Dean of Studies Software Engineering at Wilhelm Büchner University of Applied Science, Darmstadt

HOLACONF - Cloud Forward 2015: From Distributed to Complete Computing
Content

1. Introduction and Motivation
   - The RAGE Project
   - Challenges for Natural Language Processing (NLP)
   - Two important questions

2. State of the Art and Technology
   - Discovering Game Material
   - Cloud-related Technologies

3. Cloud-based Solution
   - Auto-capturing of RAGE Knowledge
   - Sufficient resources for situation specific use
   - Open issues

4. Conclusion
Introduction and Motivation

The RAGE Project

EU funded project - start: February, 2015 – end January 2019

Community Social Networks

- Software - Repository
- Media Archive
- Learn Management System
- Digital Library
- Wikis
- Learning Portals

Realizing Applied Gaming Ecosystem
Introduction and Motivation

Challenges for the Natural Language Processing

- NLP used for *capturing knowledge* from *different* and *complex object types*
- This knowledge is used for *discovering game material* and for *feeding recommender systems*
- NLP needs one *common input format*
- NLP needs *rapid elasticity*
Introduction and Motivation

Two important questions

Question 1

*How can Applied Gaming knowledge be captured and shared?*

Question 2

*How can cloud technology support this capturing and sharing?*
State of the Art and Technology

Discovering Game Material

Sharing

- Traditional IR keyword search
- Faceted Search based on classification

Capturing via NLP (with 36 gaming categories)

- Named-entity recognition (NER)
  - binary (yes/no)
  - weighted (how often)
- Text classification based on Support Vector Machines (SVM)
State of the Art and Technology
Cloud-related Technologies

Multimedia Storage and Information Retrieval
- Apache SolrCloud for *keyword and faceted Search*
- Amazon for *multimedia repository*

NLP
- GATEcloud for NER+SVM
Cloud-based Solution

Auto-capturing of RAGE Knowledge

- **Input**: Canonical Objects
  - **Output**: Extracted Features
- **Input**: Feature Vectors
  - **Output**: Classified Objects
- **Input**: Classified Objects
  - **Output**: KW+FS Service
Cloud-based Solution

Sufficient resources for situation specific use

NIST Service Model → SaaS
Deployment → Public Cloud (Amazon)
Task (supported resources) → Multimedia storage + access

NIST Service Model → PaaS
Deployment → Public Cloud (GATECloud)
Task (supported resources) → NLP and Classification

NIST Service Model → SaaS
Deployment → Public Cloud (Apache SOLR)
Task (supported resources) → Index + Classify Storage
Cloud-based Solution

Open issues

- Cloud to Cloud mediator
- Implemented as local service
- Integration of existing adaptors
- Transformed to SaaS (futur)
Conclusion

- RAGE is providing knowledge-based discovering possibilities
- Applied Gaming (AG) material need huge storage space
- Amazon cloud is used for storing all binary objects
- Knowledge needs to be captured from different RAGE object types
- For capturing, the NER+SVM cloud platform GATEcloud can be used
- The objects need to be adapted before – this implementation is a RAGE project task (at least for game software assets)
- Adapters could be implemented as SaaS (later)
- For sharing, Apache SolrCloud takes over
Towards Cloud-Based Knowledge Capturing Based on Natural Language Processing

Realizing Applied Gaming Ecosystem

Thank you very much for your attention!

Questions?

HOLACONF - Cloud Forward 2015: From Distributed to Complete Computing