NLP & Linked Data: OpeNER and NewsReader

German Rigau

http://adimen.si.ehu.eus:/~rigau

ixa

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

NAZIOARTEKO BIKAINTASUN CAMPUSA

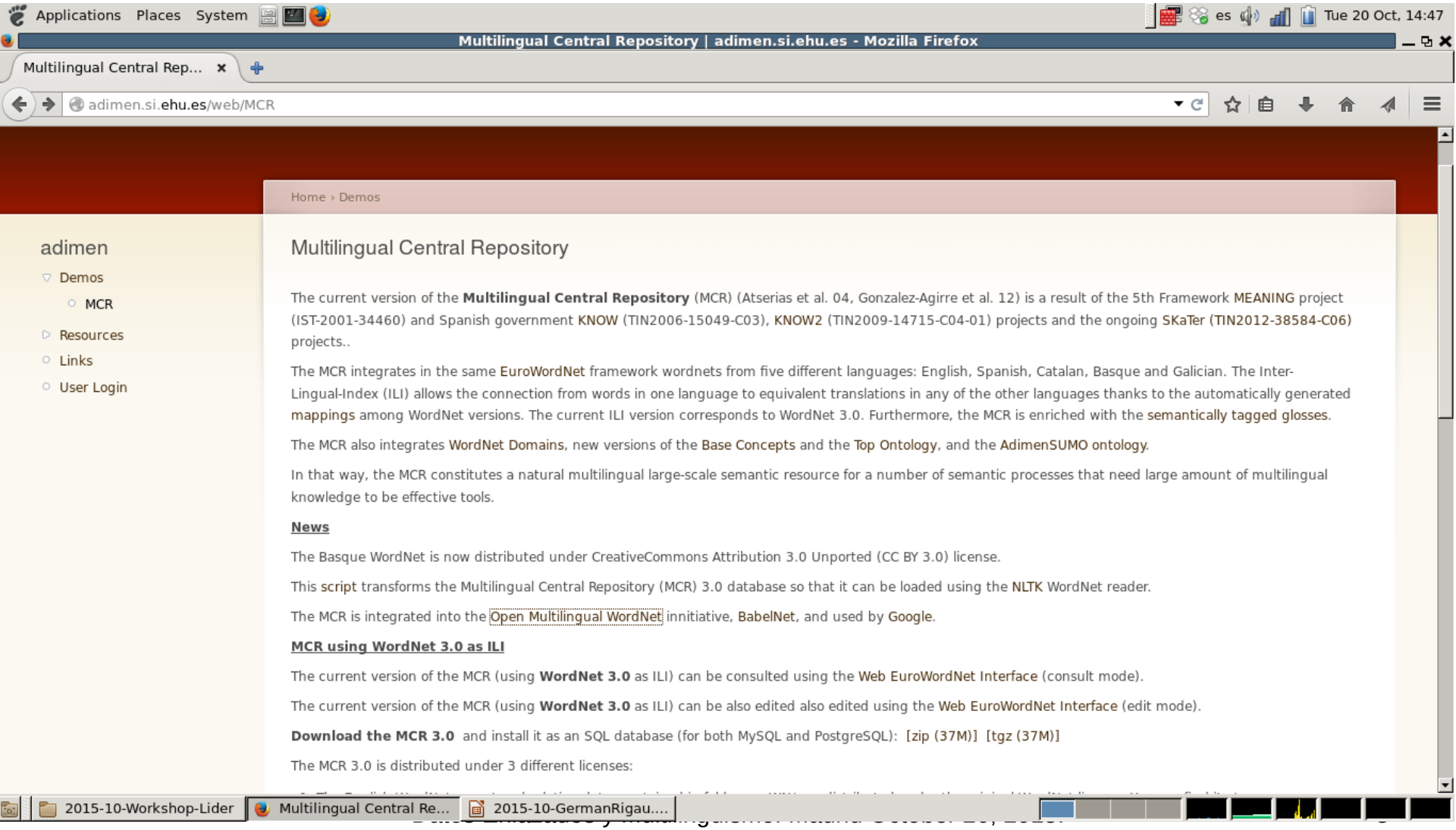
CAMPUS DE EXCELENCIA INTERNACIONAL

IXA NLP group

- **MCR** (Multilingual Central Repository)
- Academia de la lengua vasca
  - **Observatorio del léxico** (datos enlazados)
- FP7 EU research projects:
  - PATHS: Personalised Access To cultural Heritage Spaces
  - LoCloud: Local content in a Europeana cloud
  - **OpeNER**: Open Polarity Enhanced Named Entity Recognition
  - **NewsReader**: Building structured event Indexes of large volumes of financial and economic Data for Decision Making
  - Readers: Evaluation And DEvelopment of Reading System
  - QTLeap: Quality Translation by Deep Language Engineering Approaches

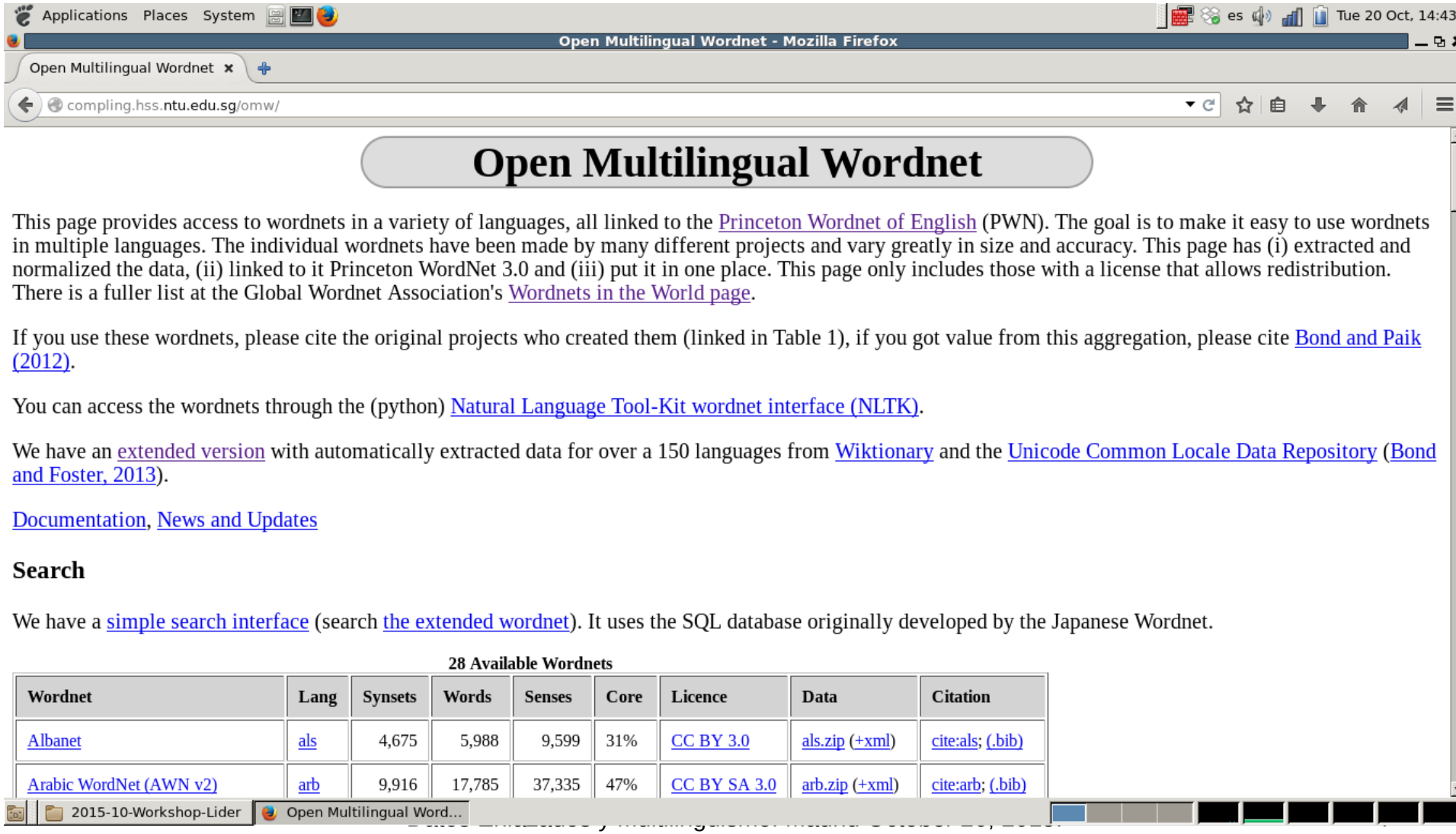
MCR

- **MCR** (Multilingual Central Repository)



MCR

- **MCR** (Multilingual Central Repository)



Observatorio del Léxico

- **Euskaltzaindia** (Academia de la Lengua Vasca)
- Corpus que se viene desarrollando desde 2007
- Euskera actual (desde 2000):
    - medios de comunicación
    - Analizado automáticamente: lematización
- Objetivo principal: monitorización del uso real del léxico
- Elhuyar, Uzei, Grupo IXA.

Observatorio del Léxico

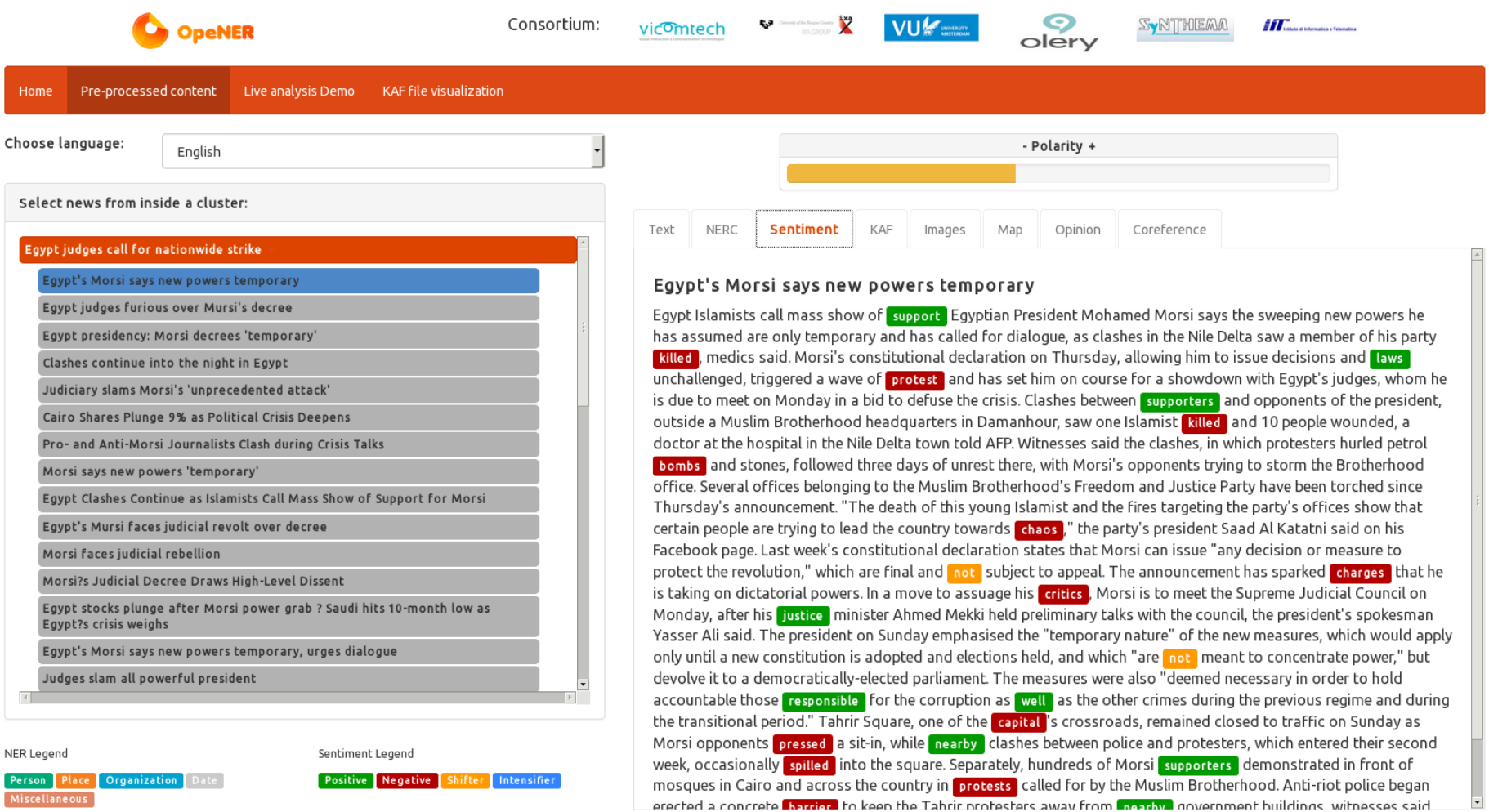
- Datos Enlazados (Corpus - Diccionario Unificado)
- Soporte para la actualización del Diccionario Unificado de la Academia (DU, *Hiztegi Batua*)
  - observar el **uso real** del léxico
  - monitorizar el cumplimiento (o no) de las normas de la Academia
- La normativa léxica dictada por la Academia ha sido integrada en la base de datos léxica
- La BD contiene también formas no estándares
- El corpus lematizado automáticamente se **enlaza al DU**
- 1 millón (2007) – 53 millones de palabras (2015)

OpeNER

- **OpeNER**: Open Polarity Enhanced Named Entity Recognition
- 2012/07/01 - 2014/06/30
- Vicomtech, VUA, CNR, Synthema, Olery, UPV/EHU.
- Main goal:
  - Easy to use NLP tools in 6 languages for big data text analytics
- Application domain:
  - Tourism
- Focusing on:
  - Opinion Mining and Sentiment Analysis of reviews
- Example:
  - Monitoring Hotel costumer opinions

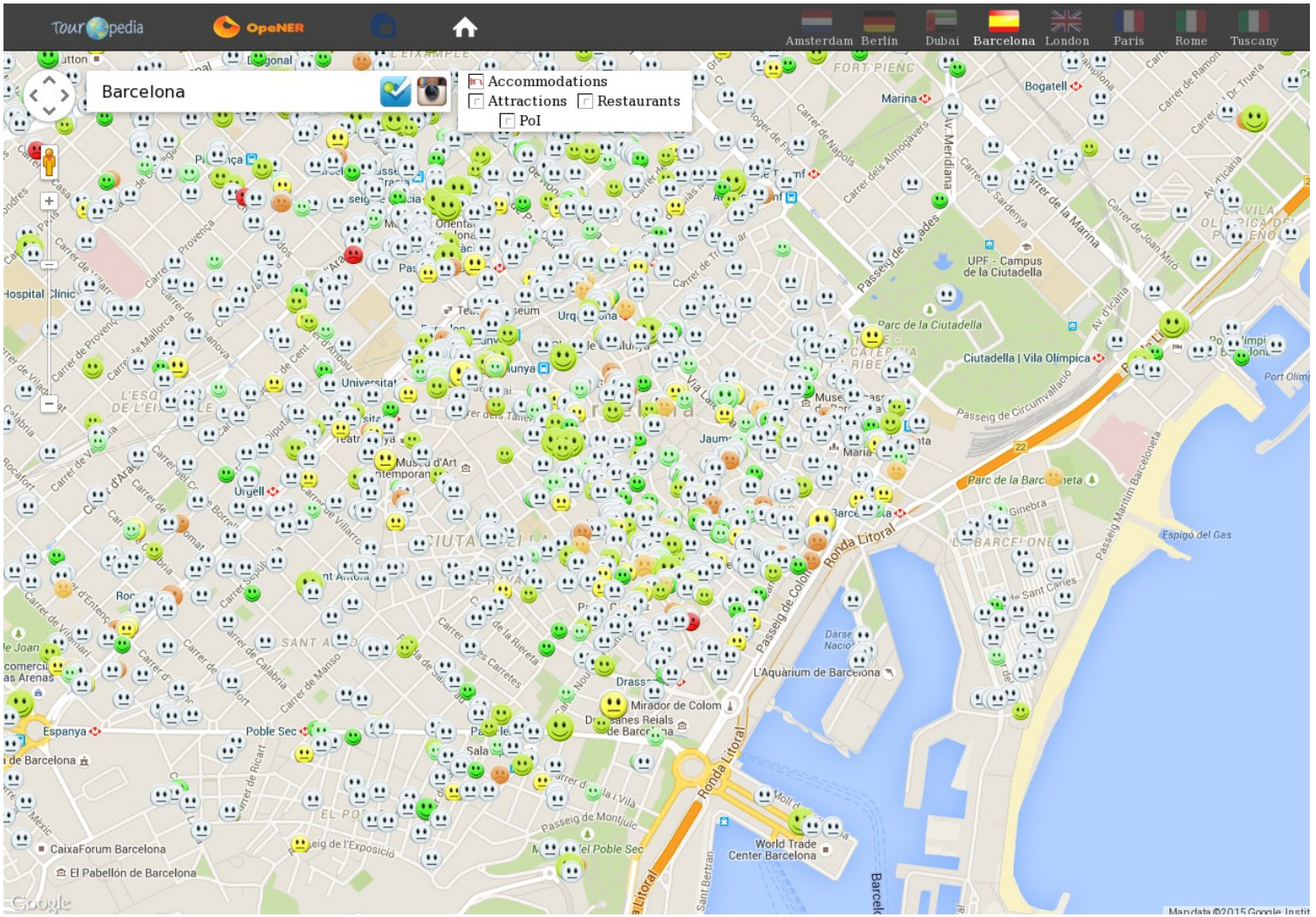
OpeNER

- **OpeNER**: Open Polarity Enhanced Named Entity Recognition



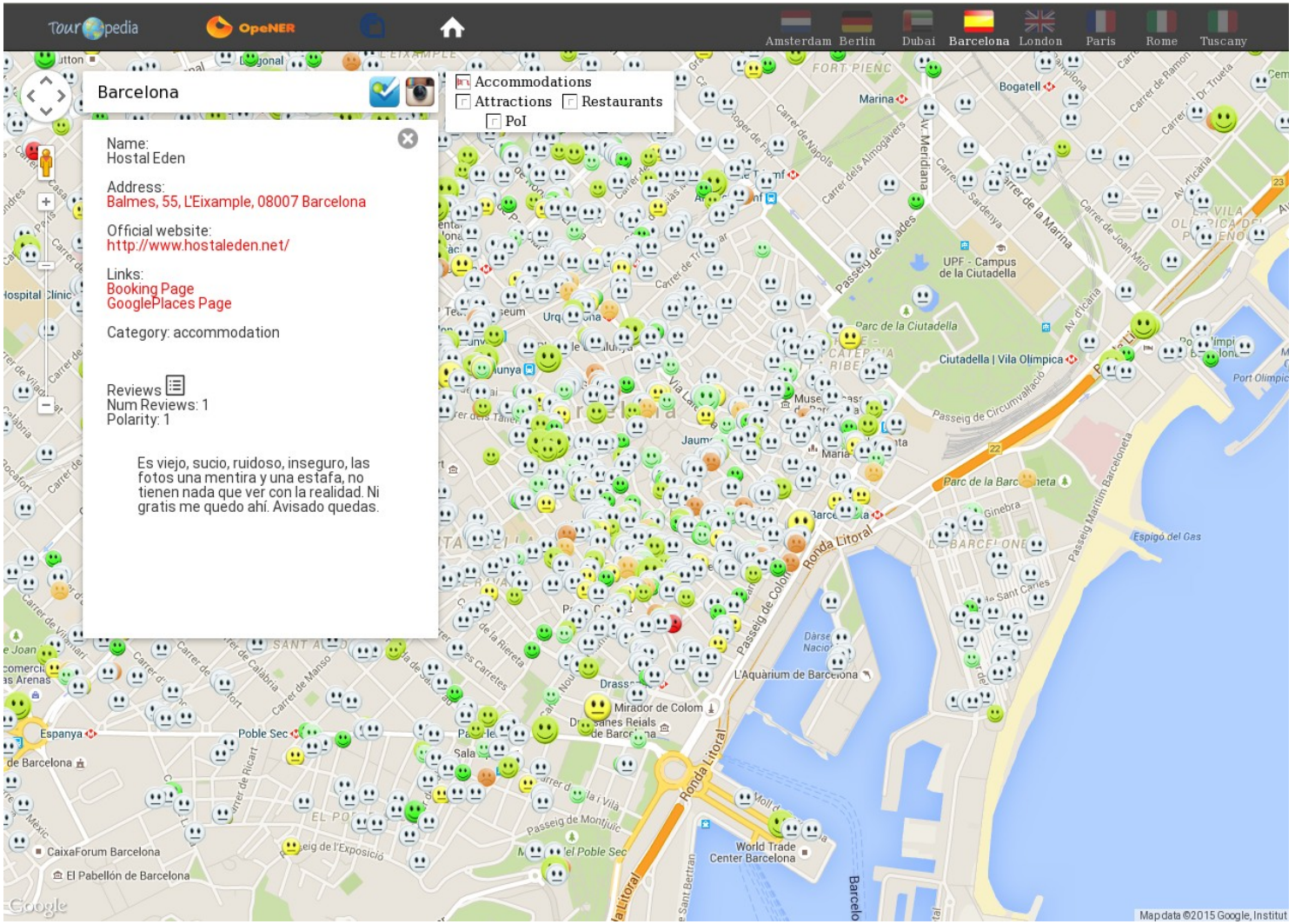
OpeNER

- **OpeNER**: Open Polarity Enhanced Named Entity Recognition



OpeNER

- **OpeNER**: Open Polarity Enhanced Named Entity Recognition

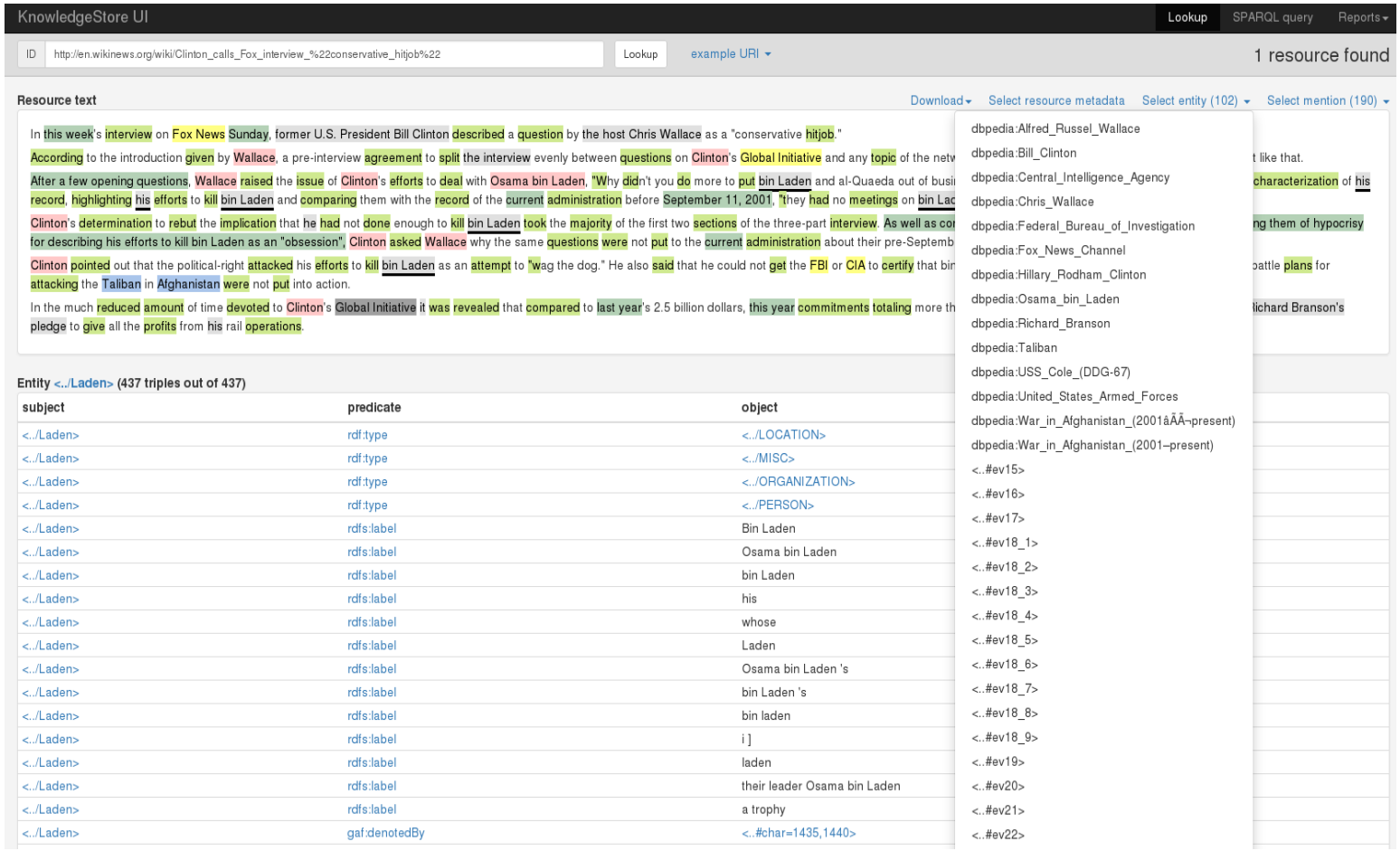


NewsReader

- **NewsReader**: Building structured event Indexes of large volumes of financial and economic Data for Decision Making
- 2013/01/01 - 2015/12/31
- VUA, FBK, LexisNexis, ScraperWiki, UPV/EHU.
- Main goal:
  - StoryLines in 4 languages at large scale
  - From NLP to Semantic Web
- Application domain:
  - Economy & Finance
- Focusing on:
  - Cross-lingual Event Mining (*who* did *what when* and *where)*
- Example:
  - Monitoring car industry, wikinews, etc.

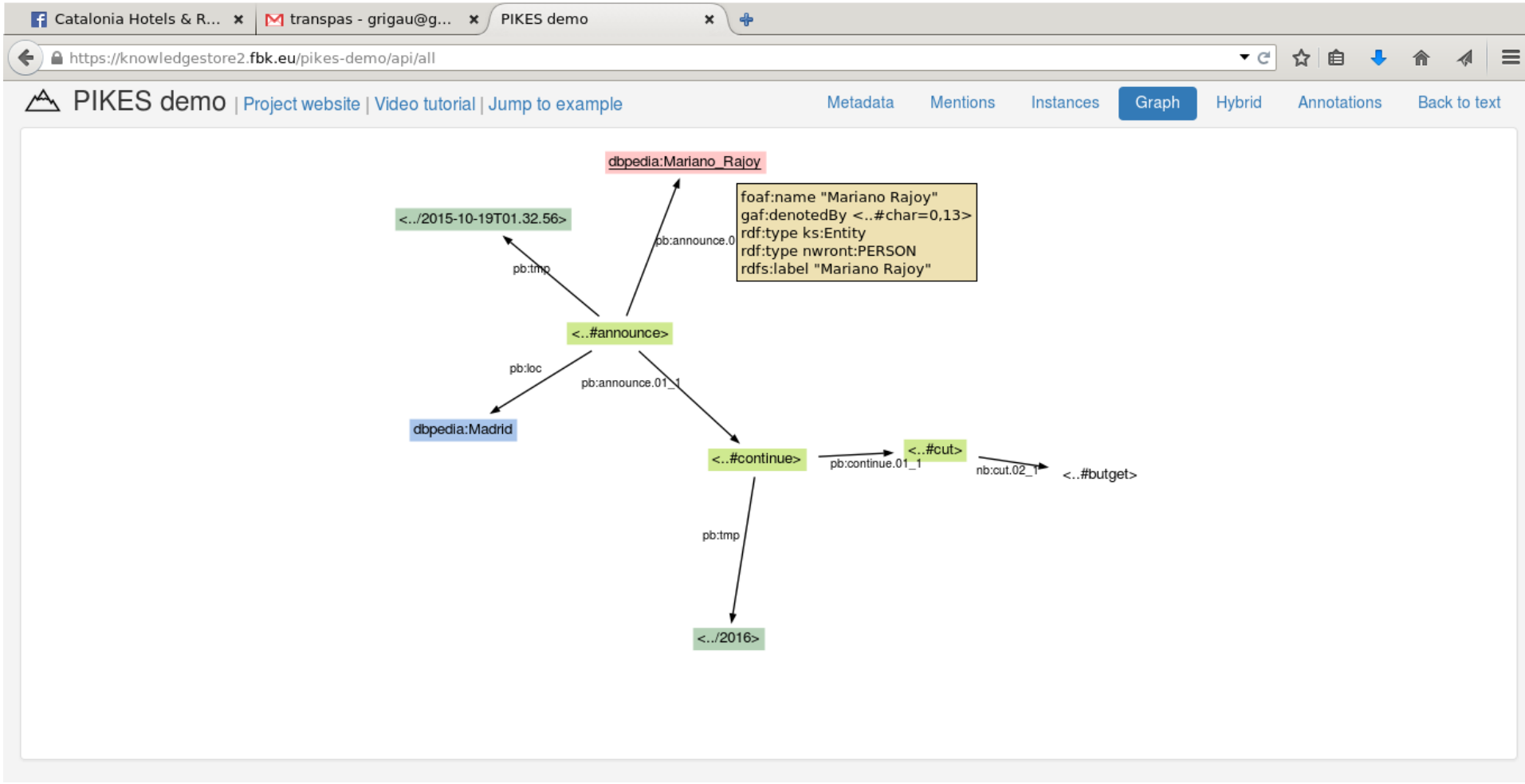
NewsReader

- **NewsReader**: Building structured event Indexes of large volumes of financial and economic Data for Decision Making



NewsReader

- **NewsReader**: Building structured event Indexes of large volumes of financial and economic Data for Decision Making



NLP & Linked Data: OpeNER and NewsReader

German Rigau

http://adimen.si.ehu.eus:/~rigau

ixa

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

NAZIOARTEKO BIKAINTASUN CAMPUSA

CAMPUS DE EXCELENCIA INTERNACIONAL