Project topics

German Rigau i Claramunt

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

IXA group

Departamento de Lenguajes y Sistemas Informáticos

UPV/EHU

Project Topics

- PT1: ukb vs. w2v on similarity/relatedness
- PT2: Mapping WordNet to Wikipedia
- PT3: New eXtended WordNet domains
- PT4: New KnowNets
- PT5: Winograd Schema Challange
- ...

ukb vs. w2v on similarity/relatedness

- UKB: http://ixa2.si.ehu.es/ukb

- word2vec:

    - https://code.google.com/archive/p/word2vec

    - http://deeplearning4j.org/word2vec

    - https://radimrehurek.com/gensim/models/word2vec.html

- Similarity/relatedness datasets:

    - http://www.cs.cmu.edu/~mfaruqui/suite.html

New KnowNets

- Current KnowNets:

- http://adimen.si.ehu.es/web/KnowNet

    - Topic Signatures (word vectors)

    - Knowledge-based WSD (SSI-Dijkstra+)

    - => KnowNets

    - Evaluation: lexical sample WSD

Mapping WordNet to Wikipedia

- Using UKB to disambiguate Wikipedia entries

- Create a sample gold-standard dataset

- Have a look at the disambiguation pages …

    - For example:

    - https://en.wikipedia.org/wiki/Party_%28disambiguation%29

    - 6 senses as a common noun in Wikipedia

    - 5 sense as a common noun in WN …

    - …

New eXtended WordNet domains

- Using WN Domains from MCR

- Use the disambiguated senses from the gloss corpus

- Propagate using ukb_ppr

- Obtain the top k synsets from the resulting vector

- Select the WN Domains from the top k synsets

- Evaluate

New KnowNets
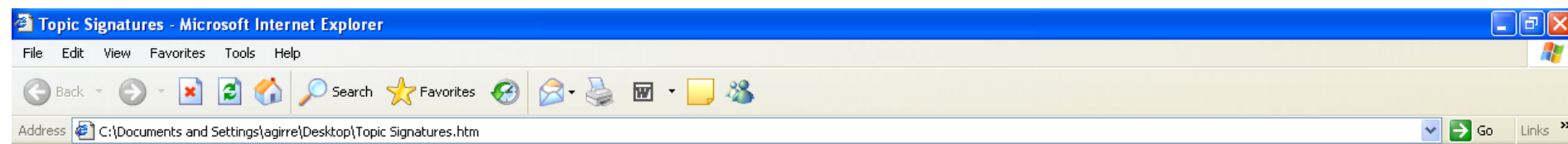
Topic Signatures Browser (all WN 1.6 *polysemous* nouns)
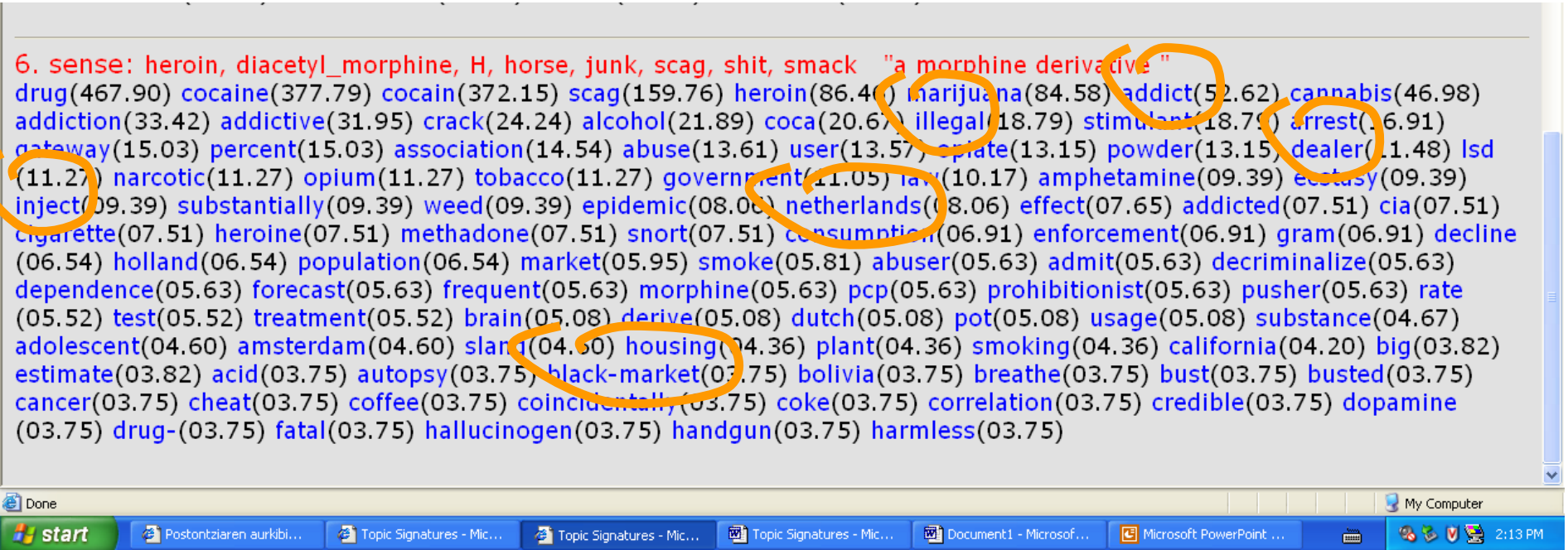
Type any noun: [horse] [submit]

horse (definitions in WordNet 1.6)
1. sense: horse, Equus_caballus "solid-hoofed herbivorous quadruped domesticated since prehistoric times "
2. sense: horse "a padded gymnastic apparatus on legs "
3. sense: cavalry, horse_cavalry, horse "troops trained to fight on horseback: "
4. sense: sawhorse, horse, sawbuck, buck "a framework for holding wood that is being sawed "
5. sense: knight, horse "a chessman in the shape of a horse's head; can move two squares horizontally and one vertically (or vice versa) "
6. sense: heroin, diacetyl_morphine, H, horse, junk, scag, shit, smack "a morphine derivative "

1. sense: horse, Equus_caballus "solid-hoofed herbivorous quadruped domesticated since prehistoric times "
polo(112.40) equus(102.66) zebra(101.61) eohippus(86.65) quagga(83.87) horse(79.18) pony(78.52) hinny(67.16) stablemate(54.63) racehorse(53.24) donkey(47.32) liver(34.45) mare(34.35) mussel(31.66) race(28.98) pinto(26.67) bangtail(26.10) workhorse(25.75) palomino(24.75) saddle(24.36) stallion(24.36) dawn(23.68) mesohippus(22.27) equid(19.48) riding(19.20) companion(18.57) harness(18.30) species(17.71) extinct(15.66) offspring(15.66) chestnut(15.61) female(15.47) hyracotherium(15.31) foal(14.61) ass(13.92) ancestor(13.22) hybrid(13.22) stable(12.67) filly(11.30) trainer(10.66) fossil(10.09) mule(10.08) thoroughbred(09.74) dreissena(08.70) breed(08.50) burro(08.35) ride(07.50) breeding(06.96) age(06.77) wild(06.62) racing(06.61) modern(06.22) champion(06.18) ago(06.05) male(05.70) broodmare(05.56) finch(05.56) mammal(05.56) dog(05.38) printer(05.38) colt(05.33) equine(05.12) owner(05.04) derby(04.87) midget(04.87) oligocene(04.87) sterile(04.87) arabian(04.69) ownership(04.69) genus(04.48) rescue(04.48) domestic(04.44) trail(04.30) eocene(04.17) mustang(04.17) subspecies(04.17) animal(03.85) bean(03.84) stud(03.84) gelding(03.82) sheep(03.82) evolution(03.63) tail(03.50) breeder(03.48) protohippus(03.48) dressage(03.41) prehistoric(03.41) rider(03.36) toe(03.23) creature(03.20) equidae(03.13) feral(03.13) sorrel(03.13) sire(03.09) mane(02.98) native(02.98) retire(02.98) evolve(02.96) tooth(02.96) cave(02.78)

New KnowNets

- Topic Signatures:
  - http://ixa3.si.ehu.es/cgi-bin/signatureak/signature_lem.cgi
  - Word vectors (acquired from the web) associated to synsets

New KnowNets

- Similar process in:
  - http://adimen.si.ehu.es/cgi-bin/WSDbyEvocation.v1/index.php
  - http://adimen.si.ehu.es/cgi-bin/WSDbyEvocation.v3/index.php

  - Using LDA on POS tagged BNC to obtain the word vectors
  - Using SSSI-Dijkstra+ for WSD

  - Try:
    - bank.n + river.n
    - bank.n + money.n

New KnowNets

- New KnowNets:

  - word2vec (word embeddings)

  - Knowledge-based WSD (SSI-Dijkstra+, UKB)

  - => New KnowNets

  - Evaluation: lexical sample WSD

New KnowNets

- New KnowNets:
  - word2vec (word embeddings)
    - https://code.google.com/p/word2vec
    - http://deeplearning4j.org/word2vec.html
    - https://radimrehurek.com/gensim/models/word2vec.html
      ...
  - Knowledge-based WSD (SSI-Dijkstra+, UKB)
    - http://ixa2.si.ehu.es/ukb
  - => New KnowNets (a very dense graph!)
  - Evaluation: lexical sample WSD
    - http://adimen.si.ehu.es/web/WSD-WN-Glosses

Winograd Schema Challange

- The trophy would not fit in the brown suitcase because **it** was too big (small). What was too big (small)?

- Answer 0: the trophy

- Answer 1: the suitcase

Winograd Schema Challange

- The bee landed on the flower because **it** had pollen.

- Answer: the flower

- The bee landed on the flower because **it** wanted pollen.

- Answer: the bee

Winograd Schema Challange

- Levesque, Hector J., Ernest Davis, and Leora Morgenstern. *The Winograd schema challenge*. Knowledge representation. 2012.

- Winograd, T. *Understanding Natural Language*. NewYork: Academic Press. 1972.

- https://www.cs.nyu.edu/davise/papers/WS.html

- http://www.hlt.utdallas.edu/~vince/data/emnlp12/train-emnlp12.txt

Winograd Schema Challange

- The bee landed on the flower because **it** had pollen.

- Answer: the flower

- The bee landed on the flower because **it** wanted pollen.

- Answer: the bee

Research topics

German Rigau i Claramunt

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

IXA group

Departamento de Lenguajes y Sistemas Informáticos

UPV/EHU