Drawbacks:

- Low hit rates
- It needs training on every tag
- Low hit rates with concrete and changing themes (like is usual in the Media)
- High number of thesaurus tags
- Few people in Documentation Departments
- It handles a high number of entities
- It could be slow and hard to design when the number of entities is high

Advantages:

- Good accuracy over general themes
- High performance (99% accuracy)
- Noise deletion
- Meaningless words discarding
- Inference ability
- Reusability

References


Acknowledgements

This work was supported by the CICYT project TIN2010-21387-C02-02 and IBERCENTRO MEDIA C.S.

Contact

General information: Ángel Luis Garrido (algarrido@heraldo.es)
Research group SID (http://sid.cps.unizar.es)