

# SQX-Lib: Developing a Semantic Query Expansion System in a Media Group

María G. Buey, Ángel Luis Garrido, Sandra Escudero,  
Raquel Trillo, Sergio Ilarri, Eduardo Mena

## Problems in relational database searches

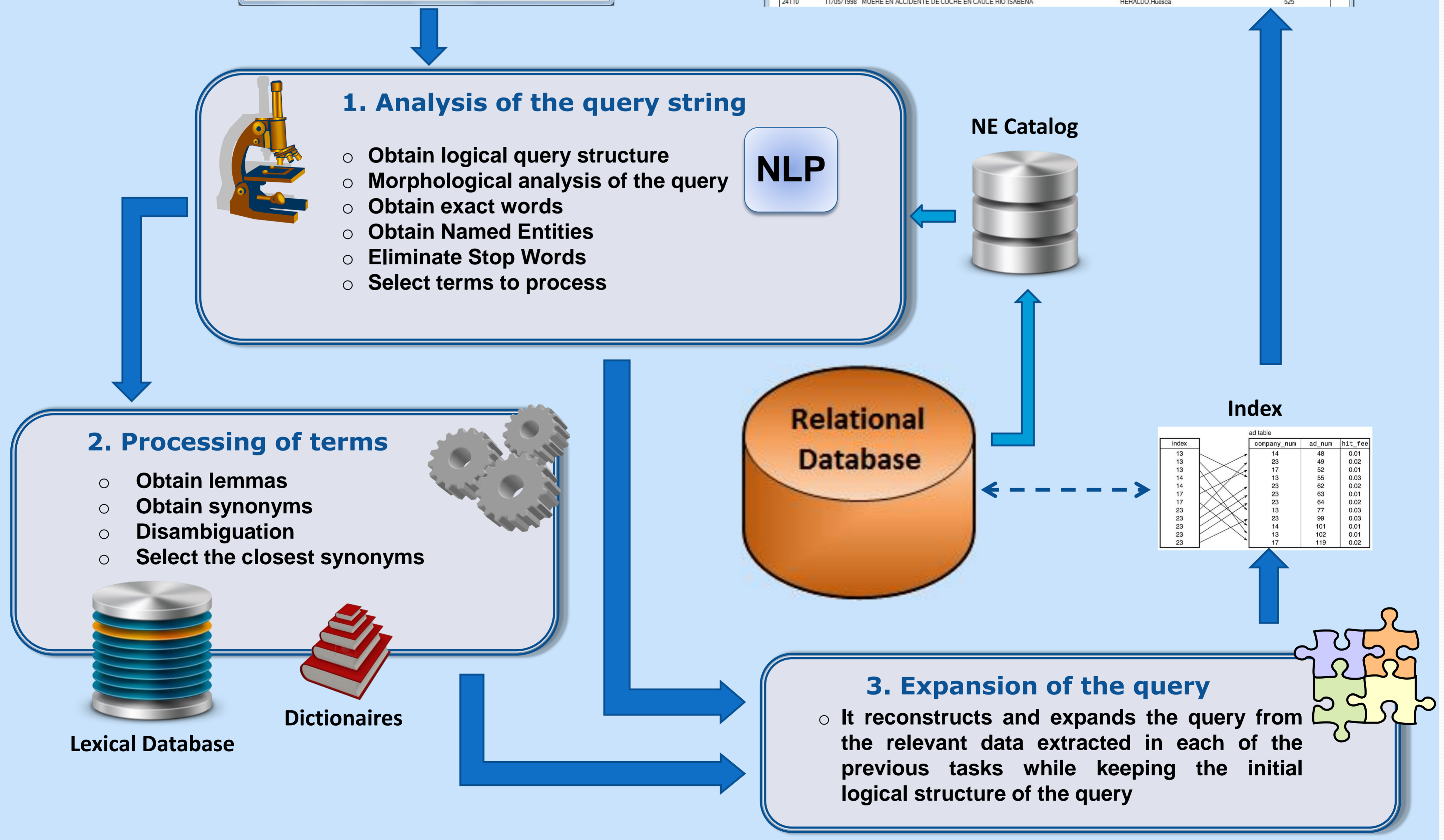
- **Term mismatch problem:** retrieving data where the user keywords exactly match with the terms of the documents. However, users often do not use the same words present in these documents. Therefore, it is needed performing a semantic analysis of the keywords to increase the effectiveness of the retrieval.
- **Low recall:** losing relevant information that the user is trying to get.
- **Low precision:** needless data that are not interesting for the user.

## Solution & Contributions

- **Automatic Query Expansion (AQE):** to expand the original query with other words that best capture the actual user intent or that simply produce a more useful query that is more likely to retrieval relevant documents.
- **SQX-Lib focuses on automatically and semantically expanding** the scope of the searches and fine-tune them by taking advantage of the Named Entities (NE) included in these queries.
- **Exploit linguistic and semantic techniques** to improve the quality of the keyword-based search process on relational repositories of the Heraldo Media Group.



ID	FECHA	TITULAR	AUTOR	NPALABRAS
1547	15/08/1991	OPERACION SALIDA RETORNO PUENTE DEL 15 DE AGOSTO	HERALDO Zaragoza	271
7147	02/03/1992	3 MUERTOS DE TOBED EN ACCIDENTE ALMONACID SIERRA	HERALDO Zaragoza	659
15734	11/05/1992	2 MUERTOS EN ACCIDENTES TRAFICO EN ARAGON	HERALDO Huesca/Zaragoza	209
19279	01/05/1993	ACCIDENTE VIA HISPANIDAD. INCENDIO ALMUNIA, FLYING	HERALDO Zaragoza	348
40135	24/08/1994	2 MUERTOS Y 2 HERIDOS EN ACCIDENTE TRAFICO A-121	HERALDO Zaragoza	371
40593	27/08/1994	2 MUERTOS TELA-SIN. 9 MUERTOS TRAFICO UNA SEMANA	AUSIETO Zaragoza	479
40765	28/08/1994	VAN 12 MUERTOS ACCIDENTES TRAFICO EN UNA SEMANA	HERALDO Zaragoza	262
41001	29/08/1994	CRONOLOGIA ACCIDENTES TRAFICO 20-27 AGOSTO 1994	HERALDO Zaragoza	259
7815	12/02/1997	MUJER MUERTA ACCIDENTE CARRETERA JALON LUMPIAQUE	HERALDO Zaragoza	226
24110	11/05/1998	MUJER EN ACCIDENTE DE COCHE EN CAUCE RIO ISABENA	HERALDO Huesca	525



## References

- [1] R. Baeza-Yates, B. Ribeiro-Neto, et al., *Modern information retrieval*, vol. 463. ACM press New York, 1999.
- [2] J. X. Yu, L. Qin, and L. Chang, *Keyword search in databases, Synthesis Lectures on Data Management*, vol. 1, no. 1, 2009.
- [3] C. Carpineto and G. Romano, *A survey of automatic query expansion in information retrieval, ACM Computing Surveys (CSUR)*, vol. 44, no. 1, 2012.
- [4] P. Vossen, *EuroWordNet: A multilingual database with lexical semantic networks*. Kluwer Academic Boston, 1998.
- [5] R. L. Cilibrasi and P. M. Vitanyi, *The Google similarity distance, IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 3, 2007.

## Acknowledgements

This work was supported by the CICYT project TIN 2010-21387-C02-02, DGA-FSE and HERALDO DE ARAGON S.A.

## Contact

General information: **María Granados Buey** ([mgbuey@unizar.es](mailto:mgbuey@unizar.es))

Research group SID (<http://sid.cps.unizar.es>)

The GENIE Project (<http://sid.cps.unizar.es/SEMANTICWEB/GENIE>)

