

# TermGen: A First Step Towards a Terminological Database in Genomics

---

Florabela Barreto<sup>1</sup>

## Abstract

The main goal of this poster is to present some aspects of the first Portuguese specialized corpus in Genomics (TermGen) and all the main steps taken for extracting the terms in this specific domain. This work rose from the necessity of filling a gap existent in the studies of terminology in Portuguese, mainly in such an important domain as this, which has been developing greatly in the last decades (cf. in vitro fertilization, organ transplants, cloning, etc.). As there are already some very advanced studies in terminology of the human genome carried by the group IULATERM at the University Pompeu Fabra in Barcelona, this will enable some future comparative work.

TermGen is a resource and knowledge database made of authentic linguistic documents of several genres, organized in an electronic format, containing, at the moment, 1.052.426 tokens. A terminological database is being built from the collection of the terms occurring in this corpus.

TermGen and the terminological database will constitute an important resource for researchers working in specialized language domains, human genome domain experts, linguistics and other professionals or students who require some kind of specialized knowledge.

---

<sup>1</sup> *e-mail*: fflloorr@gmail.com