

Resource Summary Report

Generated by ASWG on May 3, 2025

Literature-derived human gene-disease network

RRID:SCR_005653

Type: Tool

Proper Citation

Literature-derived human gene-disease network (RRID:SCR_005653)

Resource Information

URL: <http://www.dbs.ifi.lmu.de/~bundschu/LHGDN.html>

Proper Citation: Literature-derived human gene-disease network (RRID:SCR_005653)

Description: A text mining derived database with focus on extracting and classifying gene-disease associations with respect to several biomolecular conditions. It uses a machine learning based algorithm to extract semantic gene-disease relations from a textual source of interest. The semantic gene-disease relations were extracted with F-measures of 78. More specifically, the textual source utilized here originates from Entrez Gene's GeneRIF (Gene Reference Into Function) database (Mitchell, et al., 2003). LHGDN was created based on a GeneRIF version from March 31st, 2009, consisting of 414241 phrases. These phrases were further restricted to the organism Homo sapiens, which resulted in a total of 178004 phrases. We benchmark our approach on two different tasks. The first task is the identification of semantic relations between diseases and treatments. The available data set consists of manually annotated PubMed abstracts. The second task is the identification of relations between genes and diseases from a set of concise phrases, so-called GeneRIF (Gene Reference Into Function) phrases. In our experimental setting, we do not assume that the entities are given, as is often the case in previous relation extraction work. Rather the extraction of the entities is solved as a subproblem. Compared with other state-of-the-art approaches, we achieve very competitive results on both data sets. To demonstrate the scalability of our solution, we apply our approach to the complete human GeneRIF database. The resulting gene-disease network contains 34758 semantic associations between 4939 genes and 1745 diseases. The gene-disease network is publicly available as a machine-readable RDF graph. We extend the framework of Conditional Random Fields towards the annotation of semantic relations from text and apply it to the biomedical domain. Our approach is based on a rich set of textual features and achieves a performance that is competitive to leading approaches. The model is quite general and can be extended to handle arbitrary biological entities and relation types. The resulting gene-disease network shows that the GeneRIF database provides a rich knowledge source for text mining.

Abbreviations: LHGDN

Resource Type: database, data or information resource

Defining Citation: [PMID:18433469](https://pubmed.ncbi.nlm.nih.gov/18433469/)

Keywords: gene, disease, gene-disease association, text-mining, conditional random field, entity recognition

Funding: German Federal Ministry of Economics and Technology ;
THESEuropean UnionS project

Availability: Available under Creative Commons Attribution v3 Unported; please cite.

Resource Name: Literature-derived human gene-disease network

Resource ID: SCR_005653

Alternate IDs: nlx_151713

Record Creation Time: 20220129T080231+0000

Record Last Update: 20250503T055758+0000

Ratings and Alerts

No rating or validation information has been found for Literature-derived human gene-disease network.

No alerts have been found for Literature-derived human gene-disease network.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We found 1 mentions in open access literature.

Listed below are recent publications. The full list is available at [ASWG](#).

Opap K, et al. (2017) Recent advances in predicting gene-disease associations. F1000Research, 6, 578.