Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

Grado en Máster Universitario en Innovación Digital

COMPLEX DATA IN HEALTH

Studying Alzheimer's Disease

Authors:

Emanuele Alberti
Leandro Duarte
Ottavia Biagi

November 2025

Contents

Introduction

For this assignment, we were assigned to study **Alzheimer's Disease (AD)**. The following identifiers were provided: UMLS CUI `C0002395`, NCIt code `C2866`, recommended protein Apolipoprotein E (UniProt: `P02649`), transcriptome dataset `GSE300079`, and image dataset from OASIS (Kaggle).

1 Medical Terminologies and Semantic Datasets

1.1 Disease General Information

We explored biomedical databases to gather standardized information about Alzheimer's Disease.

NCBI MedGen (UID: 1853) provides the following definition: *"A degenerative disease of the brain that causes dementia, which is a gradual loss of memory, judgment, and ability to function. This disorder usually appears in people older than age 65, but less common forms of the disease appear earlier in adulthood."* The disease is classified as *Disease or Syndrome* with concept ID `C0002395`. The directly associated gene is `APP` (21q21.3), with related genes including `APOE`, `PSEN1`, `PSEN2`, `ABCA7`, `MPO`, and `PLAU`.

NCBI MeSH (ID: D000544) describes AD as a degenerative brain disease with insidious onset of dementia, impairment of memory, judgment, and attention span, followed by apraxias and global loss of cognitive abilities. Pathologically, it is marked by senile plaques, neurofibrillary tangles, and neuropil threads.

ICD-10 classifies AD under code `G30`, with subtypes: `G30.0` (early onset, before age 65), `G30.1` (late onset, after age 65), `G30.8` (other), and `G30.9` (unspecified).

Orphanet (ORPHA:1020) describes Early-onset autosomal dominant Alzheimer disease (EOAD), representing less than 1% of all AD cases, caused by mutations in `PSEN1` (69%), `APP` (13%), or `PSEN2` (2%).

Table 1 summarizes the disease codifications across vocabularies.

Table 1: Alzheimer's Disease codifications across medical vocabularies.

Vocabulary	Code
UMLS CUI	C0002395
MedGen UID	1853
MeSH	D000544
ICD-10	G30
SNOMED CT	26929004
NCIt	C2866
OMIM	104300, 516000
Orphanet	ORPHA:238616, ORPHA:1020
MONDO	MONDO:0004975
HPO	HP:0002511

1.2 SPARQL Query Results

We queried the NCIt SPARQL endpoint (`https://shared.semantics.cancer.gov/sparql`) to retrieve annotation properties for Alzheimer's Disease (code `C2866`). The query extracted the preferred label, synonyms, definition, semantic type, and UMLS CUI.

Summary of SPARQL Findings: The query retrieved annotation properties for NCIt code C2866. The disease is classified under semantic type "Mental or Behavioral Dysfunction" and maps to UMLS CUI C0002395. Seven synonyms were identified, including "Alzheimer's Dementia", "Alzheimer Disease", and "Alzheimer dementia". The definition describes AD as a progressive neurodegenerative disease characterized by nerve cell death leading to loss of cognitive function such as memory and language.