Biomedical Knowledge Graphs

Entity Types
- ● Gene
- ● Disease
- ● Drug
- ● Phenotype
- ● Pathway

Graph embeddings → Vector representations

PPARG

INS

Drug

Metformin — *treats* → Diabetes Type 2

targets

associated_with

associated_with

Diabetes Type 2

part_of

Hyperglycemia (HP:0003074)

Phenotype

has_phenotype

involves

affects

Insulin signaling pathway

Pathway

Pancreas

Tissue
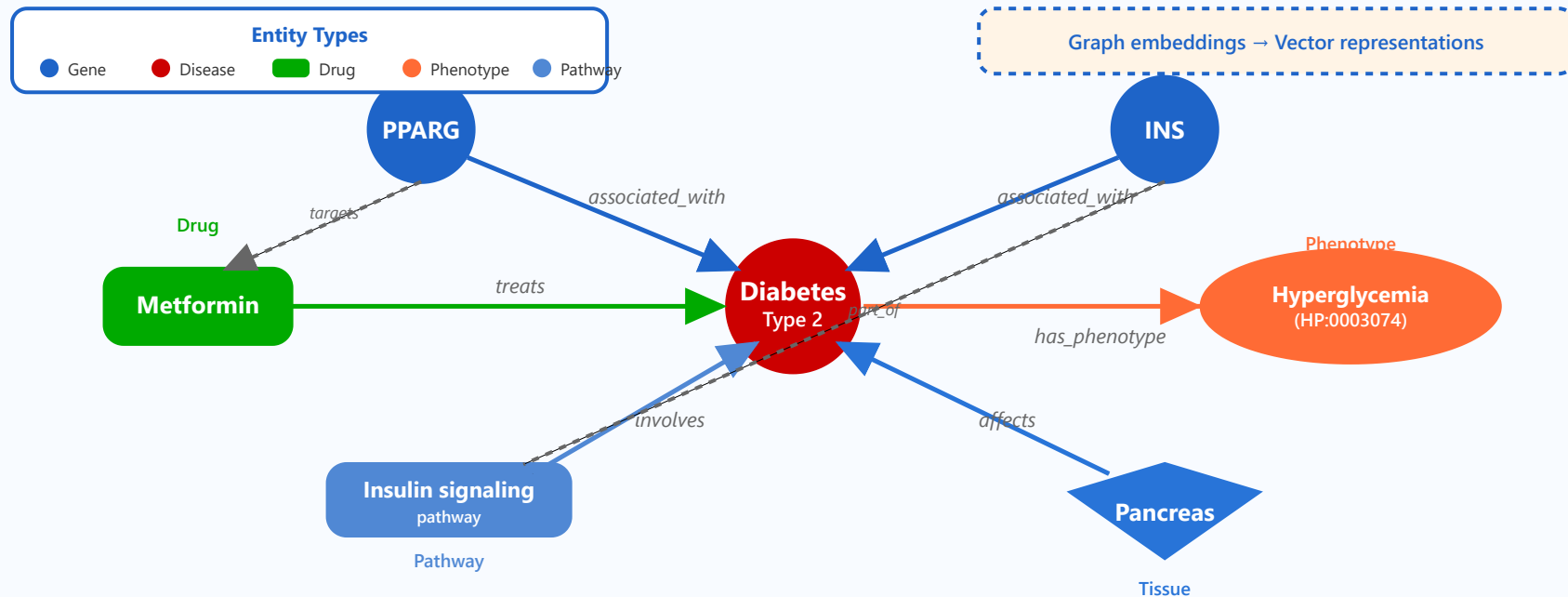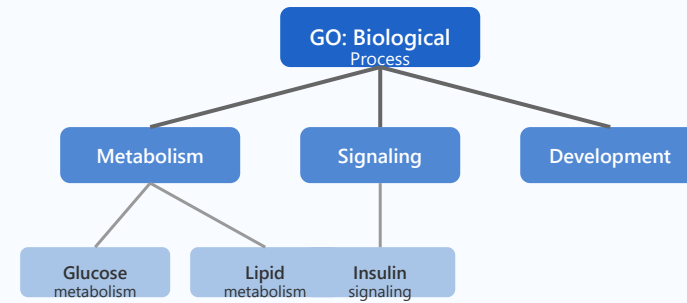
Key Components and Detailed Explanations

1 Biomedical Ontologies (생의학 온톨로지)

Biomedical ontologies are structured vocabulary systems that systematically define relationships between biological concepts. They enable consistent and standardized representation of biomedical knowledge.

Hierarchical Structure
Connected by is-a relationships

Major Ontologies:

▸ **GO (Gene Ontology)**: Gene functions, biological processes, cellular components

▸ **HPO (Human Phenotype Ontology)**: Human phenotypes and disease characteristics

▸ **DO (Disease Ontology)**: Disease classification and relationships
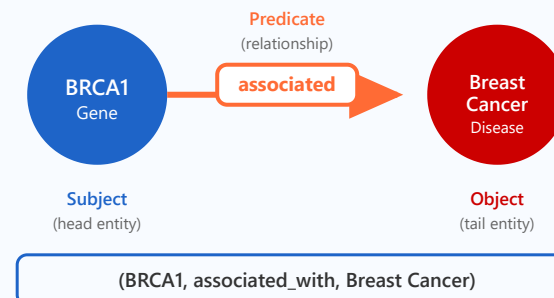
▸ **ChEBI**: Chemical entities and drug information

2 Entity Relationships (개체 관계)

The core of knowledge graphs is explicitly representing relationships between various biomedical entities (genes, diseases, drugs, phenotypes, etc.). These relationships are utilized to discover new biological insights.

(BRCA1, associated_with, Breast Cancer)

Major Relationship Types:

▸ **Gene-Disease**: BRCA1 gene mutation → Increased breast cancer risk

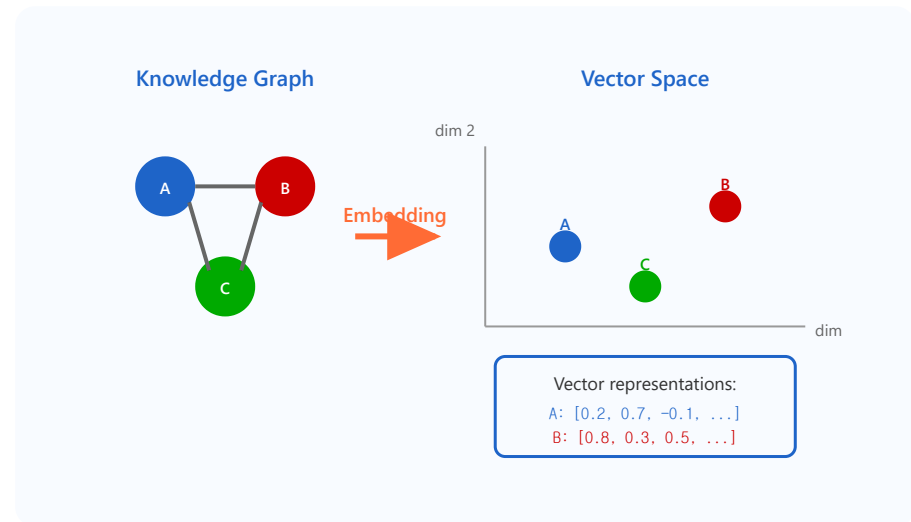▸ **Drug-Target**: Imatinib → BCR-ABL kinase inhibition

- ▸ **Disease-Phenotype**: Diabetes → Hyperglycemia
- ▸ **Gene-Pathway**: TP53 → p53 signaling pathway

③ Graph Embeddings (그래프 임베딩)

Graph embeddings transform nodes and edges of knowledge graphs into low-dimensional vector spaces. This enables machine learning models to effectively learn and utilize graph structure information.

Major Embedding Techniques:

- ▸ **TransE**: h + r ≈ t (head + relation ≈ tail)
- ▸ **Node2Vec**: Random walk-based node embeddings
- ▸ **GCN**: Graph Convolutional Networks
- ▸ **GAT**: Graph Attention Networks

Knowledge Graph → Embedding → Vector Space

Vector representations:
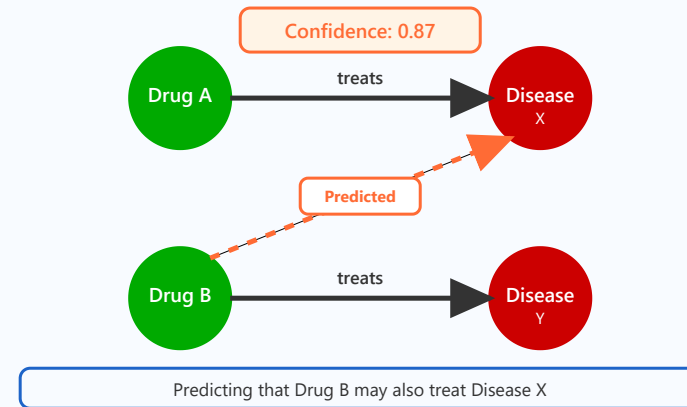A: [0.2, 0.7, −0.1, ...]
B: [0.8, 0.3, 0.5, ...]

④ Link Prediction (링크 예측)

Link prediction is a technique for predicting yet undiscovered relationships between entities in knowledge graphs. It is utilized in drug development, disease mechanism elucidation, and gene function prediction.

Application Cases:

▸ **Drug Repurposing**: Discovering new indications for existing drugs

▸ **Target Identification**: Predicting drug target proteins

▸ **Disease Gene Discovery**: Finding disease-related genes

▸ **Adverse Effect Prediction**: Predicting drug side effects

Confidence: 0.87
Drug A — treats → Disease X
Drug B — treats → Disease Y
Predicted
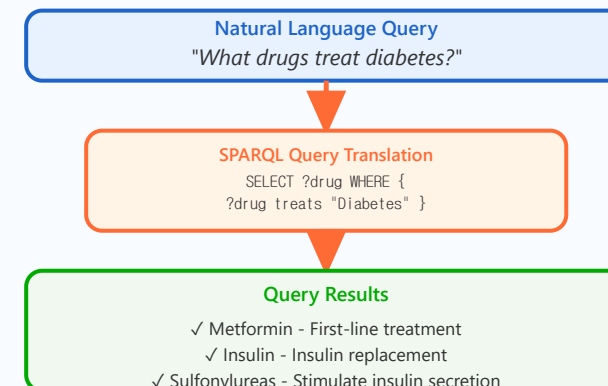Predicting that Drug B may also treat Disease X

5 Query Systems (쿼리 시스템)

Biological question answering systems allow researchers to ask complex biomedical questions in natural language and provide answers by retrieving relevant information from knowledge graphs. They utilize graph query languages like SPARQL.

Query Examples:

▸ "Find genes associated with Alzheimer's disease"

▸ "What drugs target EGFR?"

▸ "What are diabetes treatments and their side effects?"

Natural Language Query
"What drugs treat diabetes?"

SPARQL Query Translation
```
SELECT ?drug WHERE {
?drug treats "Diabetes" }
```

Query Results
✓ Metformin - First-line treatment
✓ Insulin - Insulin replacement
✓ Sulfonylureas - Stimulate insulin secretion

‣ "Which pathways does TP53 gene participate in?"