

Why knowledge graphs?

Knowledge graphs are flexible data structures which store data as nodes and ties (edges). Knowledge graphs as a general data structure are *not* limited to any particular methodology of data analysis (such as network analysis); rather, they allow the **modelling of textual data** in ways very close to original sources, enable various **reasoning capabilities** through logical inferences, and allow a **flexible schema**, which evolves with the progress of data collection and research.

CASTEMO (Computer-Assisted Semantic Text Modelling) is a **workflow for the collection of data from textual sources**. Its goal is to produce queryable, well-structured, **multi-lingual, research-oriented knowledge graphs** for **long-term and versatile use**, rather than for a single use in one publication. It is thus well-adapted, for instance, to **historical research** and projects which combine **qualitative and quantitative methodologies**.

The **CASTEMO data model** contains some pre-defined, and a large variety of flexible, user-defined features. It is based on **11 entity types (SPECTRABLOG - Statements, Persons, Events, Concepts, Territories / Texts, Resources, Actions, Living Beings, Locations, Physical Objects, and Groups)** and **three kinds of connection between entities: Properties, Relations, and References**.

Properties are flexible and extensible structures composed of an origin (the entity to which the Property is being attached), property type, and property value, and read with a “has” logic: e.g.: P current U.S. president – PROP – C area of authority – L United States of America”.

Relations, of which there are seventeen types, serve to model core ontological (e.g. Classification, Identification) and semantic (e.g. Synonymy, Superclass) connections.

References serve to relate knowledge to a specific Resource from which it has been derived.

On the whole, **knowledge graphs provide the best way of structuring complex data in ways close to the original expression in the sources**, keeping the analytical layers safely stored but neatly separated from the textual layer.

Revision #5

Created 4 March 2023 11:19:19 by Prof. Dr. David Zbírál, Ph.D.

Updated 23 June 2025 17:16:29 by Prof. Dr. David Zbírál, Ph.D.