Ontologies - Useful tools in Life Sciences and Forensics

“How today's Life Science Technologies can shape the Crime Sciences of tomorrow”

Dirk Labudde
Mittweida

04.07.2015
Mittweida
Watson vs Watson

Dr. John H. Watson

Investigator – literary character

Dr. James Dewey Watson

Cofounder of the modern biology
1952 - King’s College in London – DNA x-ray

The structure of part of a DNA double helix
DNA usually occurs as linear chromosomes in eukaryotes, and circular chromosomes in prokaryotes. The set of chromosomes in a cell makes up its genome; the human **genome** has approximately **3 billion base pairs** of DNA arranged into 46 chromosomes. The information carried by DNA is held in the sequence of pieces of DNA called **genes**. Transmission of genetic information in genes is achieved via complementary base pairing.
Transcription and translation – biological information transfer

In transcription, the codons of a gene are copied into messenger RNA by RNA polymerase. Translation in Proteins

T7 RNA polymerase (blue) producing a mRNA (green) from a DNA template (orange).
Genes and **gen products**: Proteins – functional units of living cell

Sequence – structure - function
Human Genome Project 2001


**Initial sequencing and analysis of the human genome**

The first printout of the human genome to be presented as a series of books
Modern Biology - bioinformatics

Start of the new bioinformatics in the omics era

Data (sequences)
- Storage
- Integration and organization
- Analyses
- Standardization

high throughput methods

Simulation
Visualization
Modelling
Classifying

ontologies

Genomics
Proteomics
Metabolomics
Transcriptomics

Interomics
Bibliomics

Signal transduction
Regulation of gene expression
Metabolism
System biology – part of the bioinformatics

Gen ontology on different levels
System biology – part of the bioinformatics

Protein-Protein-Interaction network

entities = proteins

Kinetics reaction

experiments predicting

Interaction:
Binding
information transfer

Drosophila Protein Interaction Map (DPiM)
data in scientific journals - literature -omics

MeSH (Medical Subject Headings) is the NLM controlled vocabulary thesaurus used for indexing articles for PubMed.

Technics of TextMining
- Analyses of the abstract content
- Ontology and semantics
background knowledge in the form of semantic networks of concept categories (called ontologies or knowledge base)
Bioinformatics - summary

Data
- Storage
- Integration and organization
- Analyses
- Standardization

**high throughput methods**

**Big data**

**Ontologies**

**Databases and query languages**

Evaluation/validation

Simulation
Visualization
Modelling
Classifying

Statistics
Energy profiles
Application - Diabetes insipidus in the context of systems biology

description of the changes in signaling pathway by point mutations based on energy profiles
Watson vs Watson

Dr. James Dewey Watson

Cofounder of the modern biology

Dr. John H. Watson

Investigator – literary character
Classical forensics and digital forensics

Locard's exchange principle

holds that the perpetrator of a crime will bring something into the crime scene and leave with something from it, and that both can be used as forensic evidence.
Our daily Life
Classical forensics and digital forensics

Physical or digital trace
- fingerprint
- DNA traces
- traces of blood and pattern
- Texts
- Short massages
- Chats
- ...

Analyses
Validation
Evaluation
circumstances of a crime
evidence

comparison of traces:
- Methods for the comparison
- Definition of similarity
- Evaluation of the results
Classical forensics and digital forensics
- With respect to bioinformatics -

bioinformatics

Information and knowledge

Model analysis

Modelling

New hypothesis

experiment

New Data
Classical forensics and digital forensics -With respect to bioinformatics -

In cases of crime often computers and other data storage media will be seized or confiscated.

Case-relevant information needs to be separated and extracted to answer and prove criminalistic questions.
Classical forensics and digital forensics
- With respect to bioinformatics -

Pre-Process
- text categorization
- separation of case-relevant files
- text extraction/OCR
- modelling of a crime ontology

Main-Process
- syntactic annotation
- semantic annotation

Post-Process
- detection of hidden semantics
Classical forensics and digital forensics
- With respect to bioinformatics -

Categorization of Forensic Texts

- bootstrapping ML combined with a set of rules
- rules determine the seed documents

Figure 6: Bootstrapping algorithm for classifying forensic texts. From the text $T_{new}$ a set of seed documents for each category is acquired using the rules annotated in the taxonomy. This set $T_{cat}$ is used to train one initial weak binary classifier per category. Subsequently, this classifier is used to classify the remaining texts $T_{remain}$ and store the new labelled documents $T_{more}$ to $T_{cat}$. Finally, the classifier is going to be improved iteratively using $T_{cat}$ until no document is left or no further improvement is possible.
Ontologies und Semantics

Why do we need ontologies?

- they define terms and symbols referring to a syntax and an association network
- prior condition for raising questions
- can support the information extraction process in different ways
- can support the visualization of results

- oriented in the Topic Map ISO-standard
- readable for humans and processable for computers
Ontologies und Semantics

Forensische Topic Map instantiation

Question Answering System
Forensische Topic Map

instantiation

Information Extraction

Template

Person

Treffen

Datum: Ort:

Person

Template Element Filling

Template Relation

Co-Reference Resolution

Named Entity Recognition

Metadaten:

author: Franz
date: 25.03.2011

Ich treffe mich morgen mit Klaus im Ritz.

...
Ontologies und Semantics

SEMANTIC TEXT ANALYZER
Ontologies und Semantics

Mona
Mobile Message ANALYZER

8 weeks later
40,000 SMS,
80 evidential SMS from 7 suspects found manually

8 hours later
20 talks containing 500 SMS incl. 80 evidential SMS from 7 suspects are extracted automatically
1-2 days manual effort

100% Precision und 67% Sensitivity
Ontologies und Semantics

Mona
Mobile Message ANALYZER

SemanTA
Semantic Text ANALYZER

AVATAR
A Victim Analysis Toolbox for Anatomic Reconstruction

SoNA
Social Network ANALYZER
Ontology and Semantics

An ontology is a specification of a conceptualization.

Definition of syntax of terms and symbols in a network of associations
Bioinformatics and Forensics - How today's Life Science Technologies can shape the Crime Sciences of tomorrow

labudde@hs-mittweida.de

www.bioforscher.de