

# Glossar 17. Mai 2010

**Applet** Ein Applet ist im Gegensatz zu den Desktopanwendungen eine Software, die nicht nur lokal auf dem Computer funktioniert, auf dem sie installiert wurde. Applets sind in einem Webbrowser laufende Java-Programme, die über eine Internetverbindung mit einem Server funktionieren.

**BorderFlow-Bibliothek** Die BorderFlow-Bibliothek implementiert einen allgemeinen Algorithmus für das Clustering der Graphen. Er verwendet ausschließlich lokale Informationen für das Clustering und erreicht ein Soft-Clustering des gegebenen Graphen.

**Clustering** Unter Clustering versteht man die Daten, basierend auf ihren Zusammenhängen, in Gruppen zu teilen. Dabei unterscheidet man zwischen Hard- und Soft-Clustering. Beim Hard-Clustering ist jedes Element in genau einem Cluster, so dass die Cluster sich nicht überlappen. Beim Soft-Clustering können Elemente auch in mehreren Clustern sein. Eine Überlappung der Cluster ist nicht ausgeschlossen.

**Data Transformation** Zur Bearbeitung der Quelldaten, die die nötigen Informationen über die Krankheiten und ihre Beziehung untereinander enthalten, müssen diese zunächst transformiert, angeordnet und aufgelistet werden. Ein weiterer Grund für die Notwendigkeit der Konvertierung der Daten aus der Quelle in ein geeignetes Zielformat ist, dass die zugrunde liegende Datei zu groß ist, um sie direkt im Graphen einzubinden.

**Desktopanwendung/ Desktopapplikation** Eine Desktopanwendung (engl. Desktopapplication) ist eine Software, die vom Benutzer auf seinem Computer lokal verwendet werden kann, um eine nützliche Funktion zu erreichen. In unserem Fall ist dies die Darstellung der Komorbiditätsgraphen nach einer Eingabe des Anwenders.

**Eingabegraphen** Der angezeigte Graph basiert auf den Eingaben des Nutzers. Der Benutzer muss Alter, Hautfarbe, die Anzahl der Stellen des ICD-9 Codes eingeben, und den ICD-Code der Krankheit, zu der er Information haben will. Aus diesen Informationen wird der benötigte Graph erstellt.

**Hudine** Hudine ist eine interaktive Website, um das ?Human Disease Network? (dt.: Netzwerk menschlicher Krankheiten) zu erforschen. Per Drag & Drop kann man dort die Zusammenhänge zwischen mehreren Krankheiten graphisch feststellen.

**ICD - International Statistical Classification of Diseases and Related Health Problems** (deutsch: Internationale statistische Klassifikation von Krankheiten und verwandter Gesundheitsprobleme) Der ICD Code ist ein Diagnoseklassifikationssystem der Medizin. Er dient zur Verschlüsselung der Diagnosen. Er kann aus einer dreistelligen Nummer bestehen oder aus einer Verschlüsselung bestehend aus vier Buchstaben und an fünfter Stelle eine Ziffer. Im Laufe der Jahre wurde der ICD-Code immer weiter entwickelt. Die derzeit gültige Ausgabe ist ICD-10. Für unser Projekt verwenden wir ICD-9CM.

**ICDToNameFile** Das ICDToNameFile ist eine Datei, in der die ICD-Codes ihren Namen zugeordnet werden.

**Indexdatei** Da die statistischen Krankheitsdaten sehr umfangreich sind, werden Indexdateien verwendet um schnell auf die Daten zugreifen zu können. Über den Schlüssel (den ICD-Code) greift man auf die Daten zu, indem der ICD-Code und ein Index auf die Zeilen im Datensatz der Quelle in der Datei gespeichert werden.

**Komorbidität** Als eine Komorbidität werden in der Medizin ein oder mehrere zusätzlich zu einer Grunderkrankung vorliegende, diagnostisch abgrenzbare Krankheits- oder Störungsbilder bezeichnet. Komorbiditäten können, müssen aber nicht ? im Sinne einer Folgeerkrankung ? ursächlich mit der Grunderkrankung zusammenhängen.

**Komorbiditätsnetzwerk** Das Komorbiditätsnetzwerk ist ein gewichteter Graph zur Darstellung der Beziehung von Krankheiten. Hierbei repräsentieren Knoten Krankheiten, die durch ihren ICD-Code angegeben werden. Das Gewicht der Kanten signalisiert die Signifikanz der Korrelation zwischen den Knoten. Fährt der Nutzer mit dem Mauszeiger über die Knoten, so erhält er Informationen wie den Namen der Krankheit. Durch Anklicken wird man zu weiteren Links geführt. Zur Darstellung werden Radial-, Simple- und Circlevew verwendet. Zwei Krankheiten stehen in Beziehung, wenn sie auffällig häufig gemeinsam, das heißt im Krankheitsverlauf einer Person, auftreten. Die Größe des Knoten ist ein Maß für die Prävalenz der Krankheit. Die Dicke der Kanten ist ein Maß für das Kantengewicht.

**Konfidenzintervall** Das Konfidenzintervall ist ein Maß für die Präzision der Lageschätzung des Relative Risk Wertes. Hierbei bezeichnet das Konfidenzintervall links die linke Grenze und das Konfidenzintervall rechts die rechte Grenze, in der man Relative Risk mit einer Wahrscheinlichkeit von 99

**Kookkurrenz** Kookkurrenz bezeichnet in unserem Fall das gemeinsame Auftreten zweier Krankheiten bei demselben Menschen. Es besteht die Annahme, dass die beiden Krankheiten voneinander abhängen, wenn sie auffällig häufig gemeinsam auftreten.

**Korrelation** Die Korrelation beschreibt die Beziehung zwischen zwei oder mehreren Krankheiten. Dies ist allerdings noch kein Hinweis darauf, ob die Krankheiten wirklich kausal voneinander abhängen.

**Pearson correlation (Phi-Korrelation)** Die Phi-Korrelation ist ein Maß für die Distanz zweier Krankheiten und dient zur Messung der Komorbidität zweier Krankheiten. Die Phi-Korrelation ist ein zuverlässiges Maß für die Komorbidität von Krankheitspaaren mit ähnlicher Prävalenz, und unterschätzt die Komorbidität zwischen seltenen und verbreiteten Krankheiten. Sie berechnet sich aus:  $(\text{Anzahl der Patienten, bei denen Krankheit 1 und Krankheit 2 diagnostiziert wurden})^2$  durch die Gesamtanzahl der Patienten  $N$ - Anzahl der Patienten, bei denen Krankheit 1 diagnostiziert wurde  $P1$  \* Anzahl der Patienten, bei denen Krankheit 2 diagnostiziert wurde  $P2$  durch  $(\sqrt{P1 * P2 * (N - P1) * (N - P2)})$

**Prävalenz** Die Prävalenz ist die Krankheitshäufigkeit. Sie wird mit einer Kennzahl der Gesundheits- und Krankheitslehre beschrieben. Sie sagt aus, wie viele Menschen einer bestimmten Gruppe an einer Krankheit erkrankt sind.

**Prefuse Library** Prefuse ist eine für Java geschriebene Programmbibliothek. Mit Hilfe der Prefuse Library lassen sich Daten modellieren, visualisieren sowie Informationen verwalten und bearbeiten. Es bietet optimierte Datenstrukturen für Tabellen, Graphiken und Bäume, eine Vielzahl von Layouts und visuelle Codierungstechniken als auch Unterstützung für Animation, dynamische Abfragen, integrierte Suche, und Datenbank-Konnektivität.

**Relative Risk** Das relative Risiko ist ein Maß für die Distanz zweier Krankheiten und dient zur Messung der Komorbidität zweier Krankheiten. Der Relative Risk Wert überschätzt den Zusammenhang zwischen selten auftretenden Krankheiten und unterschätzt die Komorbidität von häufiger auftretenden Krankheiten. Der Wert errechnet sich aus:  $(\text{Anzahl der Patienten, bei denen Krankheit 1 und Krankheit 2 diagnostiziert wurden} / \text{Gesamtanzahl der Patienten}) / ((\text{Anzahl der Patienten, bei denen Krankheit 1 diagnostiziert wurde} * \text{Anzahl der Patienten, bei denen Krankheit 2 diagnostiziert wurde}) / \text{Gesamtanzahl aller Patienten})$ .

**Signifikanzen** Die Signifikanz beschreibt den unwesentlichen Unterschied zwischen zwei Messgrößen. Wenn die Wahrscheinlichkeit, dass der Unterschied durch einen Zufall entsteht, sehr gering ist, heißen die Messgrößen signifikant.

**T-Test** Der t-Test bezeichnet eine Gruppe von Hypothesentests. Der t-Test gibt eine Angabe über die Konsistenz zweier Mittelwerte.

**View Transformation** Durch die View Transformation soll eine Änderung der Darstellung des Graphen ermöglicht werden. Durch Interaktion mit dem Benutzer kann die Darstellungsart des Graphen gewählt werden. Dabei kann er zwischen Radial, Simple und Circle View wählen.

**Viewer** Der Viewer dient zur Visualisierung des Eingabegraphen in verschiedenen Darstellungsformen: radial view, simple view, etc.. Der Benutzer hat nur Lesezugriff darauf. Er, der Viewer, ermöglicht die Exploration des Netzwerkes und basiert auf der Prefuse-Library. Zusätzlich werden Funktionen für das Zoomen und das Erhalten weiterer Informationen über die Krankheit beim Anklicken der betreffenden Knoten geboten.

**Visual mapping** Visual Mapping ist ein Prozess zur Erschließung und visuellen Darstellung der Daten. Die Abbildung der Daten in einen Graphen ermöglicht dem Benutzer schneller Zusammenhänge zu erkennen und einen besseren Überblick zu gewinnen. Der ICD-Code der einzelnen Krankheiten wird hierbei auf die Knoten des Graphen abgebildet, und die Beziehung der Krankheiten auf die Kanten. Das Kantengewicht ist ein Maß für die Signifikanz der Beziehung.