

Lastenheft

1. Zielbestimmung

Ziel der Anwendung ist es, eine Java basierte Desktop-Anwendung sowie ein abgeleitetes Applet zur Exploration von Human-Disease-Networks zu erstellen. Im Fokus stehen Komorbiditätsnetzwerke, die Prävalenz und Kookkurrenz von Krankheiten darstellen. Die Anwendung soll Graphen in verschiedenen Arten darstellen.

Zusatzinformationen aus DBpedia, Wikipedia und ICD sollen eingebettet werden, so dass der Benutzer beim Explorieren des Graphen darauf zugreifen kann.

Die Software soll den Benutzer (mit Hilfe einer integrierten Suchfunktion) in die Lage versetzen, nach verschiedenen Krankheiten zu suchen und deren Zusammenhänge mit anderen Krankheiten zu erkennen.

Des Weiteren sollen Informationen auf verschiedene Arten ausgegeben werden.

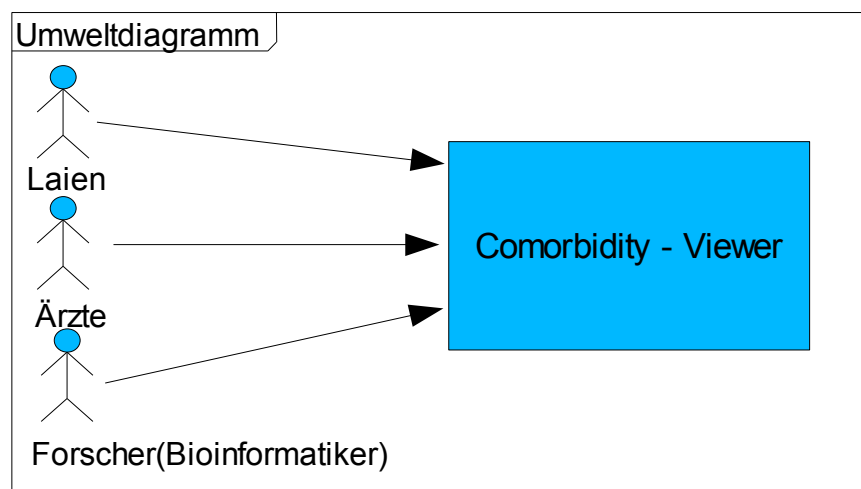
Folgende Funktionen soll die Anwendung umfassen:

- Interaktive Visualisierung des Komorbiditätsnetzwerkes als Graph
- Auswahl verschiedener Ansichten des Graphen: hierarchical, circle, spring, radial-hierarchical, simple
- Import und Benutzung von Dateien unterschiedlicher Formate (z.B. XML, CSV, Tab-Seperated)
- Suchen von Krankheiten (per Suchbegriff / ICD-9-Code) / Auswahl nach Kategorien
- Exploration des Komorbiditätsnetzwerkes (automatisch / manuell)
- Drucken des Graphen (bzw. Teilgraphen), Drucken von Informationen
- Weitere Veranschaulichung von Informationen durch Tabellen
- Links zu weiterführenden Informationen zu den Krankheiten
- Exportieren des (Teil-) Graphen als Bilddatei und im XML- / GraphML-Format

2. Produkteinsatz

Diese Software soll einerseits Ärzten oder auch Laien Zusammenhänge verschiedener Krankheiten visualisieren und andererseits zu Forschungszwecken (z.B. für Bioinformatiker) große Komorbiditätsnetzwerke darstellen und explorieren lassen.

3. Produktübersicht



4. Produktfunktionen

Muss-Funktionen:

Datenimport

/LF010/ Geschäftsprozess: **Import von ICD Tabellen**

Akteur: Nutzer

Beschreibung: Es besteht die Möglichkeit verschiedene ICD-9-Modifikationen zu verwenden. Der Datenimport wird durch ein vorher genau definiertes Format festgelegt und die jeweilige ICD-9-Version installiert.

/LF020/ Geschäftsprozess: **Indizieren und Laden von Datensätzen**

Akteur: Nutzer

Beschreibung: Es können verschiedene Datensätze für die Komorbiditäts-Statistik (z.B. von <http://hudine.neu.edu/>) im Tab-separiertem Format geladen werden. Das Format hierfür ist festgelegt bzw. vordefiniert. Beim ersten Laden, werden die Daten für einen schnelleren Zugriff indiziert.

Darstellung

/LF030/ Geschäftsprozess: **Graphansicht**

Beschreibung: Das aktuelle Teilnetz wird als Graph dargestellt inkl. optionaler Visualisierung weiterer Informationen über die Krankheiten und ihre Korrelation.

/LF031/ Geschäftsprozess: **Auswahl von Graphansichten**

Akteur: Nutzer

Beschreibung: Es kann zwischen verschiedenen Ansichten gewählt werden (hierarchical, radial-hierarchical, spring, simple, circle).

/LF032/ Geschäftsprozess: **Navigieren im Graphen**

Akteur: Nutzer

Beschreibung: In der Graphansicht kann gezoomt und gescrollt, sowie ausgewählte Knoten fokussiert werden.

/LF040/ Geschäftsprozess: **Tabellenansicht**

Beschreibung: Es besteht die Möglichkeit parallel zum Graphen eine Tabelle mit dem aktuellen Teilnetz anzeigen zu lassen. In ihr stehen die RR-, ϕ -Werte und die dazugehörigen statistischen Informationen.

/LF041/ Geschäftsprozess: **Sortieren der Tabelle**

Akteur: Nutzer

Beschreibung: Die Tabelle kann nach verschiedenen Spalten sortiert werden.

Exploration

/LF050/ Geschäftsprozess: **Suche nach Krankheiten**

Akteur: Nutzer

Beschreibung: Suche nach Krankheits- name oder -code. Ausgabe einer Liste von Übereinstimmungen.

/LF060/ Geschäftsprozess: **Expandieren / kontrahieren von Knoten**

Akteur: Nutzer

Beschreibung: Die Nachbarn eines gewählten Knoten können zum aktuellen Teilgraphen hinzugefügt oder aus ihm entfernt werden.

/LF061/ Geschäftsprozess: **Nachbarn n -ter Stufe explorieren**

Akteur: Nutzer

Beschreibung: Eine fortgeschrittene Funktion, die es dem Nutzer erlaubt, von einer Krankheit ausgehend, „Nachbarn von Nachbarn“ mit einem Mal zu explorieren. D.h. es werden nicht nur die direkten in Verbindung stehenden Krankheiten exploriert, sondern auch diejenigen, die über bis zu n weitere Krankheits-Verbindungen erreicht werden.

/LF070/ Geschäftsprozess: **Erzeugen eines Clusters um eine Krankheit**

Akteur: Nutzer

Beschreibung: Um einen gewählten Knoten kann ein Cluster erzeugt werden.

/LF071/ Geschäftsprozess: **Markieren der Cluster**

Beschreibung: Wenn ein Cluster erzeugt wurde, sind die zugehörigen Knoten identifizierbar. Nach Wunsch farblich oder durch räumliche Trennung.

/LF080/ Geschäftsprozess: **Einstellen der Darstellungsgrenzen / Wertebereiche**

Akteur: Nutzer

Beschreibung: Man kann einstellen, welche Knoten, abhängig vom Wert der Korrelation (Verbindungsstärke zwischen zwei Krankheiten), in den Teilgraphen aufgenommen werden.

Kann-/ Wunschfunktionen:

/LF100/ Geschäftsprozess: **Parametrisiertes Clustering**

Akteur: Nutzer

Beschreibung: Der Nutzer erhält die Möglichkeit zu wählen, ob mit der optimierten oder heuristischen Version des Borderflow Algorithmus gearbeitet werden soll. Außerdem können threshold und Terminationsmethode eingestellt werden.

/LF110/ Geschäftsprozess: **Speicherung der Chronik der letzten 10 Suchergebnisse**

Akteur:

Beschreibung: Möglichkeit der Speicherung der Chronik von den 10 letzten Suchergebnissen.

/LF120/ Geschäftsprozess: **Preview-Funktion für dargestellten Graph**

Akteur: Nutzer

Beschreibung: Darstellung eines Ausschnitts des Graphen als Bild.

/LF130/ Geschäftsprozess: **Drucken**

Akteur: Nutzer

Beschreibung: Drucken des Preview aus /LF120/.

/LF140/ Geschäftsprozess: **Export als Bild**

Akteur: Nutzer

Beschreibung: Exportieren des Preview aus /LF120/ als Bild.

/LF150/ Geschäftsprozess: **Export des (Teil-) Graphen**

Akteur: Nutzer

Beschreibung: Exportieren des Teilgraphen in verschiedene Formate (XML, Tab-Separiert, Datenbank).

/LF160/ Geschäftsprozess: **Import von (Teil-) Graphen**

Akteur: Nutzer

Beschreibung: Abgespeicherte Teilgraphen (z.B. im GraphML-, XML-Format) einlesen und darstellen / visualisieren → z.B. aus /LF150/ erzeugte Graphen.

/LF170/ Geschäftsprozess: **Krankheitssuche nach Kategorie**

Akteur: Nutzer

Beschreibung: Suchen nach Krankheiten mit Hilfe eines Kategoriebaums, entsprechend der ICD-9-Semantik.

5. Produktdaten

- /LD10/ ICD-9 – Daten (bis ca. 30 MB)
- /LD11/ ICD-9-CM Systematisches Verzeichnis
(zwischen 1,2 und 10 MB – je nach Datei-Format)
- /LD12/ ICD-9-CM Alphabetisches Verzeichnis
(zwischen 2 und 15 MB – je nach Datei-Format)
- /LD13/ ICD-9-GM Systematisches Verzeichnis (~ 1,5 MB)
- /LD14/ ICD-9-GM Alphabetisches Verzeichnis (~ 2,5 MB)
- /LD20/ Meta- und Indexierungsdaten für Komorbiditätsstatistik-Daten bzw. -Dateien (max. 20% der jeweiligen Dateien)
- /LD30/ History von Suchvorgängen (Chronik) und Speicherung von Programm-Einstellungen / Anwendungs-Modi (max. 2 MB)

6. Produktleistungen

- /LL10/ Import und Installation von ICD-9-Tabellen (/LF010/) muss konsistent sein bzw. Fehler (aufgrund von Unterschieden in den Länderspezifischen Modifikationen) tolerieren.
- /LL20/ Das Einbinden von Komorbiditäts-Statistikdaten (/LF020/) erfordert eine Indexierung über den geladenen Dateien. Einschließlich der Metadaten sollten maximal 20% Speicher der jeweiligen Datei ausmachen und bei der anschließenden Graphen-Darstellung für ein schnelles Auffinden und Laden der Daten sorgen (max. 1 Sek. um alle zu einem Knoten gehörigen Nachbarn in den Arbeitsspeicher zu laden).
- /LL30/ Suche nach Krankheiten per Such-Begriff (/LF050/) max. 1 Sek. für „Trefferliste“
- /LL40/ Expandieren bzw. Explorieren von Einzelknoten (/LF060/) max. 1 Sek.
- /LL41/ n-stufiges (automatisches) Explorieren des Netzwerks von einer Krankheit ausgehend (/LF061/) max. $2^n - 1$ Sek. für n Stufen
- /LL50/ Clustering (/LF070/) binnen max. 10 Sek.
- /LL60/ Schnelle Darstellung / Wechsel von Graphansichten → /LF030/ max. 2 Sek.

7. Qualitätsanforderungen

Produktqualität	Sehr gut	Gut	Normal	Nicht relevant
Funktionalität	X			
Zuverlässigkeit	X			
Benutzbarkeit		X		
Effizienz	X			
Änderbarkeit	X			
Übertragbarkeit	X			

8. Ergänzungen

- Implementierungskonzept: Es wird stets auf Teilgraphen operiert und nicht der gesamte Graph / Datensatz in den Speicher geladen → Performanz.
- Es muss noch geklärt werden, ob Spracherweiterungen des Comorbidity-Viewers zusätzlich zu der englisch sprachlichen Version realisiert werden können. Die Statistikdaten des Komorbiditätsnetzwerkes werden durch die amerikanische ICD-9-CM-Version codiert. Die Information, wie weit diese Codes, bezogen auf die reinen Krankheiten, mit ICD-9-Modifikationen anderer Staaten übereinstimmen, bekommen wir (hoffentlich) demnächst von der DIMDI (Deutsches Institut für medizinische Dokumentation und Information). Diese Frage betrifft die Funktion /LF010/.