

Projekt AGB-10

Pflichtenheft

Verantwortliche: Pamela Gelfert

Gruppe: SWP10-9

17. Mai 2010

Inhaltsverzeichnis

1 Zielbestimmungen	4
1.1 Musskriterien	4
1.2 Wunschkriterien	5
1.3 Abgrenzungskriterien	5
2 Produkteinsatz	5
2.1 Anwendungsbereiche	5
2.2 Zielgruppen	6
2.3 Betriebsbedingungen	6
3 Produktübersicht	7
4 Produktfunktionen	7
4.1 Dateien importieren	7
4.2 Visualisierung der Daten	7
4.3 Exploration des Graphen/Funktionalität zur Erkundung des Graphen . . .	9
4.4 Suchfunktionen	10
5 Produktdaten	10
6 Produktleistungen	12
7 Qualitätsanforderungen	13
8 Benutzungsoberfläche	13
9 Nichtfunktionale Anforderungen	14
10 Technische Produktumgebung	14
10.1 Software	14
10.2 Hardware	14
10.3 Orgware	14
11 spezielle Anforderungen an die Entwicklungsumgebung	15
11.1 Software	15
11.2 Hardware	15
11.3 Orgware	15

12 Gliederung in Teilprodukte	15
13 Ergänzungen	16

1 Zielbestimmungen

Das Ziel des Softwareprojektes Graph-Browser ist die Entwicklung einer Desktopanwendung und eines dazugehörigen Applets zur Darstellung und Erkundung der Prävalenz und Kookkurrenz von Krankheiten in Komorbiditätsnetzwerken. Die Erkundung des Graphen soll durch verschiedene Funktionen erleichtert werden.

1.1 Musskriterien

- Darstellung von Komorbiditätsnetzwerken, die Knoten repräsentieren hierbei die ICD-Codes der Krankheiten, die Kanten sind ein Maß für die Signifikanz der Korrelation der Krankheiten
- Anwendung steht sowohl als Applet, als auch als Desktopanwendung zur Verfügung
- Import verschiedener Dateitypen im tab-separierten Format
- Verwendung importierter Dateien für die Erstellung des Graphen
- Eingabe des ICD-Codes in eine integrierte Suchfunktion
- Ausgabe der Suchergebnisse in einer Tabelle
- Spezifikation der Suche durch die Wahl des Geschlechtes und der Hautfarbe
- Einblenden von Knoten und Kantenbeschriftung
- Verwendung unterschiedlicher Graphtypen zur Darstellung der Komorbiditätsnetzwerke (simple view, radial, circle, hierarchical radial und hierarchical view)
- Clustern des Graphen
- Veranschaulichung der Information von Kookkurrenz und Prävalenz einer Krankheit in eine Tabelle
- Einbettung von Links zu weiterführenden Informationsseiten zu den Krankheiten im Graphen
- Erforschung weiterer Zusammenhänge durch Anklicken neuer Knoten
- Exploration des Eingabegraphen durch Zoomen, Scrollen und Fokussieren

1.2 Wunschkriterien

- Eingabe der Krankheit in die integrierte Suchfunktion
- Einblendung einer Beschriftung der Knoten mit Namen
- Einbettung der Links sowohl im Graphen, als auch in der Tabelle
- Vorschau auf externe Internetseiten

1.3 Abgrenzungskriterien

- Dateien, die nicht im tab-separierten Format vorliegen, können nicht zur Grapherstellung verwendet werden.
- Die Software ist nicht fähig Daten zu interpretieren. Dementsprechend müssen importierte Dateien in ihrem Aufbau dem Datensatz von Hudine gleichen (siehe /D10/).

2 Produkteinsatz

2.1 Anwendungsbereiche

Die Software dient zur Erkennung von Zusammenhängen zwischen Krankheiten. Sie soll helfen möglichst gezielt Vorkehrungen gegen bestimmte Krankheiten zu treffen. Deswegen ist sie besonders geeignet für eine Anwendung in Krankenhäusern oder Arztpraxen. Desweiteren erleichtert sie das Erkennen von Mustern und Zusammenhängen zwischen Krankheiten. Deshalb kann sie in Forschungseinrichtungen oder an Universitäten genutzt werden, um zum Beispiel Statistiken aufzustellen.

Generell kann die Software in jedem privaten Haushalt genutzt werden. Die graphische Anschauung vereinfacht das Erkennen von Zusammenhängen und ist einfacher zu überblicken, als Statistiken und Tabellen.

2.2 Zielgruppen

Die Zielgruppen sind unter anderem Ärzte, Mediziner und Homöopathen, die vor allen bei Krankheiten, die häufig miteinander auftreten, versuchen können das Ausbrechen der korrelierten Krankheit zu verhindern.

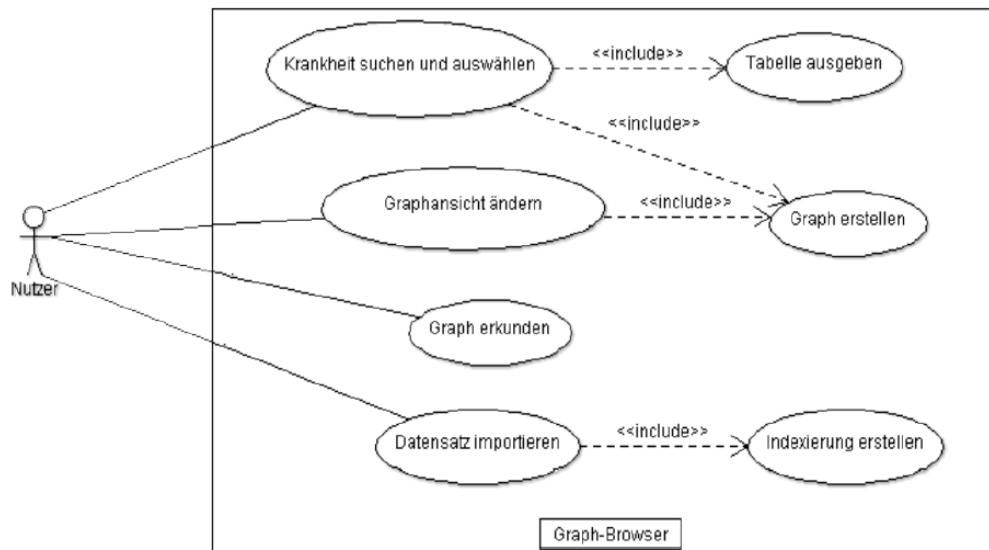
Ebenso dient die Software Forschern, Bioninformatikern und eventuell auch Studenten. Doch auch die Allgemeinheit kann den Graph-Browser nutzen, um sich auf diesem Gebiet zu informieren. Der Datensatz von Hudine erfasst hauptsächlich die Korrelation von Krankheiten von älteren Personen und ist deswegen am aussagekräftigsten für die ältere Bevölkerung. Allerdings beinhaltet der Datensatz viele Infektionskrankheiten und Krankheiten bzw. Beschwerden, die mit einer Schwangerschaft verbunden sind, nicht. Die Software ist folglich nicht für daran leidende Personen ausgelegt, sofern der Datensatz von Hudine verwendet wird.

2.3 Betriebsbedingungen

Der Graph-Browser steht jedem Computer zur Verfügung. Um auf weiterführende Informationsseiten im Internet zu gelangen oder das Applet zu nutzen, ist eine Internetverbindung nötig. Allerdings kann die Software auch ohne Internet als Desktopapplikation genutzt werden.

Die Anwendung ist für einen wartungsfreien, ständigen Betrieb ausgelegt.

3 Produktübersicht



4 Produktfunktionen

4.1 Dateien importieren

/F010/ Geschäftsprozess: Laden von Dateien und Indexdatei erstellen

Akteur: Nutzer

Beschreibung: Falls der Nutzer für die Erstellung des Graphen nicht den Standarddatensatz von Hudine verwenden möchte, so kann er einen anderen Datensatz im tab-separierten Format hochladen. Für diesen Datensatz legt das System eine Indexdatei an, die ein schnelles Navigieren über die eingelesenen Daten ermöglicht.

4.2 Visualisierung der Daten

/F20/ Geschäftsprozess: Darstellung des Graphen

Akteur: System

Beschreibung: Auf Basis der Krankheitsdaten wird der Graph zu einer vom Nutzer gewählten Krankheit erstellt. Die Krankheiten werden als ICD-Code in den Knoten dargestellt.

Die Kantengewichte entsprechen den Signifikanzen der Korrelationen zwischen den

Krankheiten. Dargestellt werden die Graphen nach dem im Punkt /F40/ ausgewählten Stil.

/F30/ Geschäftsprozess: Auswahl eines Graphentypen zur Darstellung

Akteur: Nutzer

Beschreibung: Der Nutzer kann über einen Button zwischen verschiedenen Typen von Graphen (simple view, radial, circle, hierarchical radial und hierarchical view) zur Darstellung der Signifikanz der Korrelationen zwischen den Krankheiten wählen.

/F40/ Geschäftsprozess: Einstellen eines Wertes für das Kantengewicht

Akteur: Nutzer

Beschreibung: Der Nutzer kann entscheiden, ab welchem Kantengewicht Nachbarn angezeigt werden sollen. Als Standardwert wird für Phi-Korrelation = 0,1 und für Relativ Risk = 5 angenommen.

/F50/ Geschäftsprozess: Einblenden von Knotenbeschriftungen

Akteur: Nutzer

Beschreibung: Der Nutzer kann wählen, ob im Graphen Informationen wie Krankheitsnamen oder ICD-Codes mit angezeigt werden sollen.

/F60/ Geschäftsprozess: Zusatzinformationen anzeigen

Akteur: System

Beschreibung: Nach der Auswahl eines Knotens werden in einem Popup-Fenster Zusatzinformationen angezeigt. Dies sind Krankheitsname, ICD-Code, Prävalenz und Links zu weiteren Informationsseiten zu diesen Krankheiten. Berührt der Nutzer mit dem Mauszeiger eine Kante, wird das entsprechende Kantengewicht eingeblendet.

/F70/ Geschäftsprozess: Auswahl zwischen Relativ Risk und Phi-Korrelation als Kantengewicht für den Graphen

Akteur: Nutzer

Beschreibung: Der Nutzer kann entscheiden, ob die Signifikanz der Korrelationen zwischen den Knoten durch Phi-Korrelation oder Relativ Risk dargestellt werden soll.

/FW80/ Geschäftsprozess: Zusätzlichen Knoten (Krankheit) zur Anzeige hinzufügen

Akteur: Nutzer

Beschreibung: Wenn ein Knoten, aufgrund seiner geringen Korrelation zum gewählten bzw. gesuchten Knoten nicht angezeigt wird, kann der Benutzer diesen Knoten optional anzeigen lassen (durch Auswahl des Namen oder ICD-Codes aus der Tabelle von /F100/).

/F90/ Geschäftsprozess: Erstellen einer Tabelle

Akteur: System

Beschreibung: Zusätzlich zum Graphen wird eine Tabelle mit den Krankheiten angezeigt, welche mit der gewählten Krankheit korrelieren. Diese beinhaltet Informationen wie Krankheitsname, ICD-Code, Prävalenz und Kantengewicht.

/FW100/ Geschäftsprozess: Sortierung der Tabelle

Akteur: Nutzer

Beschreibung: Die zum Graphen erstellte Tabelle kann nach verschiedenen Werten (Krankheitsname, ICD-Code oder Kantengewicht) sortiert werden.

4.3 Exploration des Graphen/Funktionalität zur Erkundung des Graphen

/F110/ Geschäftsprozess: Clustern des Graphen

Akteur: Nutzer

Beschreibung: Auf Wunsch des Nutzers kann der Graph geclustert werden. Dabei werden die Knoten in Abhängigkeit vom Kantengewicht zu Gruppen zusammengefasst.

/F120/ Geschäftsprozess: Funktionalität für die Exploration der Eingabedaten

Akteur: Nutzer

Beschreibung: Dem Nutzer stehen Funktionen wie scrollen, zoomen und fokussieren für die Erkundung des Graphen zur Verfügung.

/FW130/ Geschäftsprozess: Auswahl/Fokussierung rückgängig machen

Akteur: Nutzer

Beschreibung: Es wird ein Button bereitgestellt, mit dem der Nutzer zum vorher fokussierten Knoten zurückspringen kann, bis der Ursprungsknoten seiner Exploration erreicht ist.

/FW140/ Geschäftsprozess: Vorschau auf externe Informationsseite anzeigen

Akteur: Nutzer

Beschreibung: Wenn der Nutzer mit dem Cursor über einen Link fährt, wird eine kleine Vorschau auf die Informationsseite angezeigt

4.4 Suchfunktionen

/FW150/ Geschäftsprozess: Suchen und Markieren eines Knotens im aktuellen Graphen

Akteur: Nutzer

Beschreibung: Um bei großer Knotenanzahl im aktuellen Graphen gesuchte Knoten schneller zu finden, kann im Suchfeld ein Krankheitsname eingegeben werden. Ein oder mehrere passende Knoten werden dann graphisch hervorgehoben.

/F160/ Geschäftsprozess: Suchen einer Krankheit

Akteur: Nutzer

Beschreibung: Der Nutzer gibt entweder den Namen einer Krankheit oder den ICD-Code ein. Daraufhin wird eine Liste mit Übereinstimmungen ausgegeben, welche den ICD-Code, die Krankheitshäufigkeit in Prozent und den kompletten Namen der Krankheit enthält.

/F170/ Geschäftsprozess: Spezifikation der Suche

Akteur: Nutzer

Beschreibung: Der Nutzer kann neben der Auswahl einer Krankheit ein Geschlecht und eine Hautfarbe (schwarz oder weiß) angeben. Dadurch spezialisieren sich die statistischen Ergebnisse (Kookkurrenz, Prävalenz).

/F180/ Geschäftsprozess: Generierung einer Liste mit Übereinstimmungen

Akteur: Nutzer

Beschreibung: Für den gesuchten Knoten wird eine Liste von Krankheiten erzeugt, die der Nutzer gemeint haben könnte.

/F190/ Geschäftsprozess: Auswahl aus der Liste mit Übereinstimmung

Akteur: Nutzer

Beschreibung: Aus der Liste /F190/ wählt der Nutzer per Mausklick einen Krankheitsnamen oder einen ICD-Code aus und spezifiziert somit die Krankheit, auf deren Grundlage der Graph dann erstellt wird.

5 Produktdaten

/D10/ Quelldatensatz von Hudine (alle Patienten (13039018 Patienten)):

- ICD9 CM-Code für Krankheit 1
- ICD9 CM-Code für Krankheit 2

- Prävalenz von Krankheit 1 (angegeben in : Anzahl der daran erkrankten Personen)
- Prävalenz von Krankheit 2 (angegeben in : Anzahl der daran erkrankten Personen)
- Kookkurrenz von Krankheit 1 und Krankheit 2 (angegeben in : Anzahl der erkrankten Personen)
- Relatives Risiko in Prozent
- Relatives Risiko 99% Konfidenzintervall (links) in Prozent
- Relatives Risiko 99% Konfidenzintervall (rechts) in Prozent
- Phi-Korrelation
- T-Test Wert

6088553 Tupel

/D20/ Quelldatensatz von Hudine (Frauen (7598529 Patienten)):

Aufbau: siehe /D010/

4882376 TupelIndexFile über die übergebenen Dateien

/D30/ Quelldatensatz von Hudine (Männer (5440490 Patienten):

Aufbau: siehe /D010/

3980759 Tupel

/D40/ Quelldatensatz von Hudine (weiß (7598529 Patienten)):

Aufbau: siehe /D010/

4871245 Tupel

/D50/ Quelldatensatz von Hudine (schwarz(5440490 Patienten)):

Aufbau: siehe /D010/

2007916 Tupel

/D60/ Quelldatensatz von Hudine (männlich :weiß (4910632 Patienten)):

Aufbau: siehe /D010/

3775421 Tupel

/D70/ Quelldatensatz von Hudine (männlich :schwarz (386663 Patienten)):

Aufbau: siehe /D010/

1239552 Tupel

/D80/ Quelldatensatz von Hudine (weiblich :weiß (6835054 Patienten)):

Aufbau: siehe /D010/
4660064 Tupel

/D90/ Quelldatensatz von Hudine (weiblich schwarz(596432)):

Aufbau: siehe /D010/
1557867 Tupel

/D100/ IndexFile über die übergebenen Dateien

- ICD9-CM Code
- Links zu den benötigten Zeilen im Datensatz
- Links zu Webseiten, die Informationen über die jeweilige Krankheit enthalten

/D110/ ICDToNameFile, das den ICD-Codes ihre Namen zuweist

- ICD Code
- Name zum ICD Code

6 Produktleistungen

/L10/ Die Erstellung des Graphen soll nur wenige Sekunden dauern.

/L20/ Die Umwandlung der verschiedenen Graphen darf nur wenige Sekunden dauern.

/L30/ Die Suche nach Übereinstimmungen für einen Suchbegriff darf nur wenige Sekunden dauern.

/L40/ Das Erweitern des Graphen um neue Knoten darf nur wenige Sekunden dauern.

/L50/ Das Hochladen und Einlesen beliebig großer Dateien muss möglich sein.

/L60/ Das Clustern (/F110/) darf nur wenige Sekunden dauern.

/L70/ Das Erstellen der Indexdatei darf nicht zu lange dauern

/L80/ Die Indexdatei darf höchstens 25% der Größe des ursprünglichen Datensatzes haben

/L80/ Die Umrechnung in Prozent (zum Beispiel bei dem Wert Prävalenz) wird auf drei Stellen nach dem Komma gerundet.

7 Qualitätsanforderungen

Produktqualität	sehr gut	gut	normal	nicht relevant
Funktionalität				
Angemessenheit	X			
Richtigkeit	X			
Interoperabilität			X	
Ordnungsmäßigkeit		X		
Sicherheit	X			
Zuverlässigkeit				
Reife	X			
Fehlertoleranz		X		
Wiederherstellbarkeit			X	
Benutzbarkeit				
Verständlichkeit	X			
Erlernbarkeit	X			
Bedienbarkeit	X			
Effizienz				
Zeitverhalten		X		
Verbrauchsverhalten			X	
Änderbarkeit				
Analysierbarkeit	X			
Modifizierbarkeit	X			
Stabilität	X			
Prüfbarkeit	X			
Übertragbarkeit				
Anpassbarkeit		X		
Installierbarkeit		X		
Konformität		X		
Austauschbarkeit		X		

8 Benutzungsoberfläche

/B10/ Die Bedienung der Benutzeroberfläche erfolgt über die Maus (eine reine Tastaturbedienung wird nicht unterstützt).

/B20/ Eingaben werden über die Tastatur getätigt.

/B30/ Das Bildschirmlayout lehnt sich stark an die Oberfläche von <http://hudine.neu.edu/> an.

9 Nichtfunktionale Anforderungen

- Der Datensatz von Hudine darf nicht umgeschrieben werden. Der Benutzer kann ihn sich zwar ansehen, allerdings nicht schreibend darauf zugreifen.
- Zu jeder Klasse des Javaprogrammes für das Softwareprojekt Graph-Brwoser sollen eine aussagekräftige und verständliche JavaDoc-Kommentare verfasst werden. Die Kommentierung richtet sich nach den Richtlinien des Qualitätssicherungskonzeptes.
- Zum besseren Umgang und Einarbeitung mit dem Programm wird ein Benutzerhandbuch und eine Designbeschreibung angefertigt

10 Technische Produktumgebung

Das Produkt läuft auf einem Arbeitsplatzrechner mit graphischer Benutzungsoberfläche.

10.1 Software

- Beliebiges Betriebssystem
- Webbrowser mit JavaScript Fähigkeit (für das Applet)

10.2 Hardware

- PC mit Monitor, Tastatur und Maus

10.3 Orgware

- Optional: Internetverbindung (um weitere Informationseiten zu besuchen)

11 spezielle Anforderungen an die Entwicklungsumgebung

11.1 Software

- Eclipse
- JavaCompiler
- Texteditor
- LaTeX
- Beliebiges Betriebssystem

11.2 Hardware

- es wird die gleiche Hardware wie in 10.2 genannt benötigt

11.3 Orgware

- Aufgabenbeschreibung auf OLAT
- E-Mail oder ICQ für die Kommunikation der Gruppe
- Subversion (SVN) für das Austauschen und bearbeiten der Dokumente

12 Gliederung in Teilprodukte

In diesem Projekt gibt es keine Gliederung in Teilprodukte.

13 Ergänzungen

- Das Produkt sollte gut erweiterbar sein, um später zusätzliche Funktionen, wie weitere Ansichten oder Auswertungskriterien auf die Daten, einbinden zu können.
- Im zu Grunde liegenden Datensatz wird ICD-9-CM (International Classification of Diseases, Clinical Modification) verwendet. Der ICD-Code ist eine an klinische Bedürfnisse angepasste Version. Diese Codierung besteht aus fünf Stellen.
- Durch die Struktur von Prefuse ist bereits eine hervorragende Übertragbarkeit gewährleistet.
- Für die Zuordnung der Namen zu den ICD-Codes wird die Seite:
<http://people.dbmi.columbia.edu/hripcsak/icd9/1tabular.html>
verwendet.