

INFRASTRUKTURKONZEPT FÜR DIE INTEGRIERTE VERSORGUNG – MPI-IMPLEMENTIERUNG

Lamprecht M¹, Thiele S¹, Häber A¹

Kurzfassung

Durch die Einführung von sektorübergreifenden Versorgungsformen wie der integrierten Versorgung werden immer komplexere Anforderungen an die Kommunikationsinfrastruktur gestellt. Im folgenden Beitrag wird eine mögliche Infrastruktur unter Verwendung eines zentralen Master Patient Index (MPI) vorgestellt. In heterogenen Anwendungssystem-Landschaften findet ein Austausch von patientenbezogenen Daten über Integrationsplattformen statt. Die Probleme liegen vor allem in der Datenintegrität (insbesondere Dublikaterkennung). Mit diesem Fokus wurden Algorithmen innerhalb der Integrationsplattform Ensemble entwickelt und umgesetzt.

Abstract

New intersectoral forms of medical care (e.g. integrated care) have complex requirements to the it-infrastructure. The following article introduces a possible infrastructure using a central Master-Patient-Index (MPI). In order to the heterogeneous it-infrastructures the patient related data need to be exchanged across an integrationplatform. The dataintegrity and the detection of duplicates are problems in this context. To solve these problems algorithms for the detection of duplicates are developed and implemented with the integration platform Ensemble.

Keywords – Master Patient Index (MPI), record linkage, duplicate detection, integrated care

1. Einleitung

Mit dem Gesundheitsmodernisierungsgesetz 2004 hat der Gesetzgeber im § 140a ff SGB V die schon seit mehreren Jahren geforderte und nur im Einzelnen gelebte sektorübergreifende Versorgung mit dem Begriff der Integrierten Versorgung für das deutsche Gesundheitswesen festgeschrieben [3]. Krankenkassen erhalten im § 140a ff SGB V durch Verträge die Möglichkeit, ihren Versicherten eine abgestimmte Versorgung anzubieten, bei der Haus- und Fachärzte, ärztliche und nichtärztliche Leistungserbringer, ambulanter und stationärer Bereich sowie gegebenenfalls Apotheken koordiniert zusammenwirken. Krankheiten, wie beispielsweise chronische Erkrankungen, die häufig eine Behandlung über mehrere Sektoren erfordern, sollen nicht abgegrenzt in heterogenen Einrichtungen behandelt werden. Die Abläufe, die zur Genesung beitragen, sollen abgestimmt über Sektorengrenzen hinweg erfolgen. Um das zu ermöglichen, fordert der § 140a die gemeinsame Dokumentation der Leistungserbringer über Einrichtungsgrenzen hinweg. Zur Vereinfachung der Kommunikation bieten sich hier elektronische Lösungen an.

¹ Westsächsische Hochschule Zwickau

Dazu ist es erforderlich, dass die Anwendungssysteme der Leistungserbringer miteinander kommunizieren können. Betrachtet man die in den Krankenhäusern und im niedergelassenen Bereich installierten Softwareprodukte, so stellt man fest, dass eine gemeinsame Dokumentation nicht unproblematisch ist.

Eine sektorübergreifende Dokumentation oder intersektorale Kommunikation erfordert umfassende Anpassungen an der Dokumentation und Kommunikation in beiden Bereichen.

1. 1. Intersektorale Kommunikation

Unter Intersektoraler Kommunikation versteht man eine Kommunikationsbeziehung zwischen den Sektoren im Gesundheitswesen. Wenn diese aus Zeit- und Kostengründen elektronisch ablaufen soll, müssen die vorherrschenden IT-Strukturen mit ihren Anwendungssystemen beachten werden. Da im deutschen Gesundheitswesen Sektoren, wie ambulanter und stationärer Sektor, ihre eigenen speziellen Anforderungen besitzen, werden in den verschiedenen Sektoren auch unterschiedliche Anwendungssysteme benutzt, die sich unabhängig voneinander entwickelt haben (Abbildung 1).

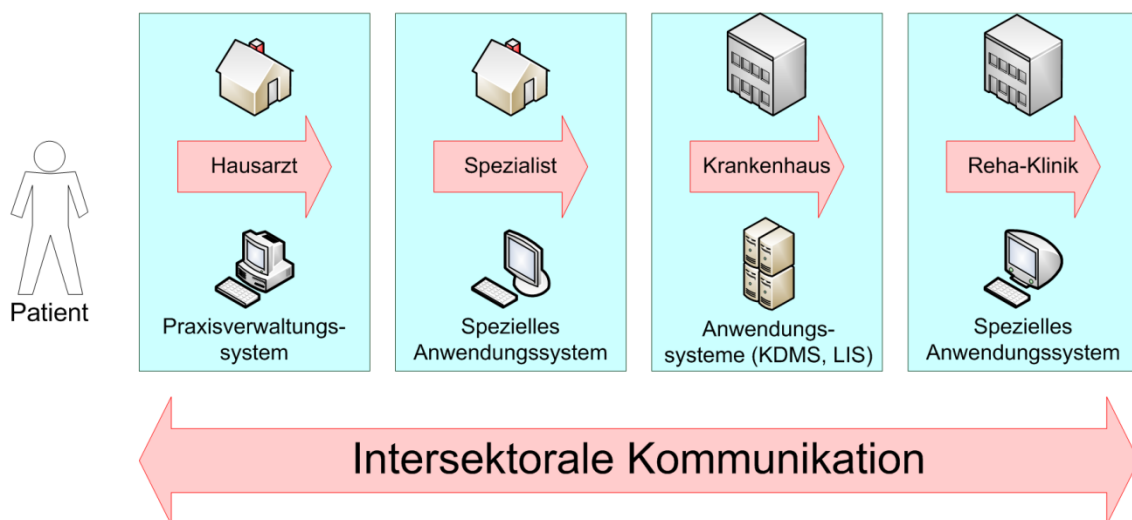


Abbildung 1: Intersektorale Kommunikation zwischen verschiedenen Einrichtungen des Gesundheitswesens

1. 2. Anwendungssysteme der Sektoren

Die Informationssysteme in den Krankenhäusern sind in der Regel sehr heterogen aufgebaut. Hier werden unterschiedliche Anwendungssysteme für verschiedene Dokumentationsaufgaben verwendet (z.B. Patientenverwaltungssystem für die Aufnahme, Entlassung und Abrechnung der Patienten, Klinisches Dokumentations- und Managementsystem für die dokumentarischen Aufgaben der Ärzte und Pflegekräfte insbesondere auf den Stationen, Laborinformationssystem für die Gerätesteuerung und Befundschreibung im klinisch chemischen Labor). Von den unterschiedlichen medizinischen Fachgebieten (Innere Medizin, Chirurgie, Intensivmedizin) werden ganz unterschiedliche Anforderungen an die Informationstechnik (Software und Hardwarekomponenten) gestellt. Die Anwendungssysteme im Krankenhaus kommunizieren in der Regel unter Verwendung des Standards Health Level 7 (HL7) miteinander. Für den Datenaustausch zwischen den einzelnen Anwendungssystemen des Krankenhauses wird ein Kommunikationsserver zur Übersetzung und Nachrichtenweiterleitung benötigt.

Der niedergelassene Bereich setzt vorwiegend Praxisverwaltungssysteme (PVS) ein, die eine chronologische, nicht fallbezogene Dokumentation beinhalten und den Arzt bei der Administration seiner Praxis als Unternehmen unterstützt (Kassen-, Privat-, und BG-Abrechnung etc.). Außerdem enthält ein Praxisverwaltungssystem ein Managementsystem zur zentralen Verwaltung von Patienten- und Behandlungsfalldaten und ein Arztarbeitsplatzsystem für die Falldokumentation. Die Anforderungen an die Hardware sind eher minimal. Eine PVS Software wird häufig auf einem Desktop PC installiert. Die Schnittstellen zum Datenex- und -import sind in der Regel sehr einfach gehalten und unterstützen die von der Kassenärztlichen Bundesvereinigung vorgegebenen Datenformate, z. B. zur Übertragung von Abrechnungsinformationen.

Diese unterschiedlich vorherrschende Infrastruktur in den einzelnen Sektoren muss bei einer umzusetzenden Integrationslösung beachtet werden.

2. Methoden

2.1. Eindeutige Patientenidentifikation mit Hilfe eines MPI

Beim sektorübergreifenden Datenaustausch müssen die innerhalb der Institutionen verwendeten unterschiedlichen Patientenidentifikationen zu beachten, die von Einrichtung zu Einrichtung variieren. Damit wird ein Patient in den verschiedenen Einrichtungen über unterschiedliche, nur intern gültige Nummern identifiziert. Bei einer intersektoralen Kommunikation muss gewährleistet sein, dass die versendeten Daten eines Patienten beim Empfänger wieder eindeutig zuzuordnen sind.

Um dieses zu erreichen, gibt es mehrere Lösungsansätze. Bei nur zwei gekoppelten Anwendungssystemen könnte jedes System die Identifikationsdaten des anderen in einer sogenannten Mappingtabelle speichern. Das ist bei einer steigenden Anzahl anzubindender Systeme nicht mehr wartbar. Wollen mehrere Systeme, beispielsweise innerhalb eines Ärztenetzes, Daten austauschen, ist ein Master-Patient-Index (MPI) ein Lösungsansatz. Der Begriff Master-Patient-Index wird einerseits verwendet für eine eindeutige ID, unter der die Patientenidentifikationen der angeschlossenen Einrichtungen zusammengeführt und gespeichert werden [9]. Der Referenzpatient mit dieser globalen ID enthält dann alle demographischen Identifikationsdaten zu einem Patienten. Des Weiteren wird die Bezeichnung MPI als Sammelbegriff für Datenbank- und Anwendungsfunktionalität verwendet, um die Identifikationsdaten zu speichern und zu administrieren.

Als zentrale Datenbank kann der MPI für Abfragen verwendet werden, um bei einem Datenaustausch die Patientenidentifikationsnummer (PID) einer angeschlossenen Einrichtung zu erfahren. So ist ein Patient bei einem Datenaustausch eindeutig identifizierbar und Datenverlust wird vermieden. Damit es in der MPI Datenbank nicht zu Mehrfachregistrierungen kommt, muss der MPI die Funktionalität des Clearings, d.h. der Dublettenerkennung und Zusammenführung (Mergeing) unterstützen [2].

2.2. Ähnlichkeitssuche in Patientendatenbanken

Da Patienten Identitäten in unterschiedlichen Anwendungssystemen letztendlich den gleichen Patienten repräsentieren können, muss der MPI die Duplettenerkennung unterstützen. Dadurch kann eine eindeutige Weiterleitung von Patientendaten ermöglicht werden.

Die Identifizierung von Duplikaten innerhalb verteilter Patientendatenbanken kann algorithmisch durch den Sorted Neighborhood Algorithmus [8] realisiert. Hierzu werden Datensätze mit einem frei wählbaren Datensortieralgorithmus sortiert mit dem Ziel, dass potentielle Duplikate möglichst

nahe beieinander liegen und somit in einer Nachbarschaft existieren. Anschließend müssen die Datensätze miteinander verglichen werden, die in einer Nachbarschaft liegen. Hieraus resultiert eine Gruppe (Cluster) vermutlicher Duplikate.

Das Verfahren der Sortierten Nachbarschaft stützt sich grundsätzlich auf drei Schritte. Es wird davon ausgegangen, dass die zu sortierenden Daten in einer Tabelle vorliegen. Es seien neben Nachname und Vorname weitere demografische Daten wie Telefonnummer, Anschrift etc. erfasst.

Im ersten Schritt wird aus relevanten Teilen von Tabellenattributen der Schlüssel erzeugt. Der Schlüssel existiert lediglich „virtuell“ und ist nicht eindeutig, um ihn zur Sortierung verwenden zu können. Ist die Schlüsselerzeugung abgeschlossen, wird im zweiten Schritt des Verfahrens mit der Sortierung begonnen. Die Sortierung erfolgt lexikografisch nach dem erzeugten Schlüssel. Die gängigsten Datenbanksysteme bringen bereits vorimplementierte Sortieralgorithmen mit, welche direkt zum Einsatz kommen können. Die Sortierung ordnet dann die Datentupel so an, dass wahrscheinliche Duplikate möglichst nah beieinander liegen.

Die Überprüfung für das Mergeing wird anhand eines vordefinierten Fensters (sliding window) mit vorher festgelegter Größe über die Datensätze geschoben. Es werden nur Datensätze verglichen, die innerhalb dieses Fensters liegen, der Vergleich findet paarweise statt. Ein Fenster mit festgelegter Größe w ist mit der Anzahl N der Tupel definiert als $2 \leq w \leq N$. Für den Vergleich kommen verschiedene Distanzmaße zum Einsatz, beispielsweise Levenshtein-, Damerau-Levenshtein- und Jaro-Winkler-Distanz [1,6,7].

2.3. Zielstellungen

Um Anforderungen, die bei einer Vernetzung im Rahmen einer Integrierten Versorgung von mehreren Anwendungssystemen zu erfüllen sind, zu untersuchen und einen Lösungsansatz für einen MPI zu bieten, wurde eine Testumgebung aufgebaut. Anhand dieser Testumgebung sollten zum einen die Möglichkeiten der Kommunikation mit ausgewählten Anwendungssystemen getestet werden: Welche Schnittstellen besitzt das Anwendungssystem und welche Standards unterstützt das jeweilige System für den Datenaustausch.

Außerdem sollte für die MPI Komponente untersucht werden, wie eine algorithmische Unterstützung bei einer Duplettenerkennung realisiert werden kann.

2.4. Aufbau einer Testumgebung

Für die genannte Testumgebung wurden folgende Anwendungssysteme aus dem Gesundheitswesen verwendet:

- das Praxisverwaltungssystem DOCconcept der DOCexpert Gruppe,
- das klinische Dokumentations- und Managementsystem MCC der Meierhofer AG sowie
- das klinische Dokumentations- und Managementsystem i.s.h.med der Siemens GSD mit dem Patientenverwaltungssystem IS-H der SAP AG.

Um den Aufwand für Nachrichtenaustausch und Schnittstellenanpassungen so gering wie möglich zu halten, wurde als Middleware die Integrationsplattform Ensemble der Firma Intersystems zur Kopplung der Anwendungssysteme verwendet [5].

Ensemble ist ein Kommunikationsserver, der im Gesundheitswesen vor allem zur Nachrichtenweiterleitung und –transformation zum Einsatz kommt. Im Hintergrund arbeitet die Datenbank Caché. Diese wird genutzt um geroutete Nachrichten abzuspeichern und Nachrichtenschemas wie beispielsweise HL7 in verschiedenen Versionen abzurufen. Die hinterlegten Nachrichtenschemen werden für eine versionsgenaue Nachrichtenverarbeitung benötigt. Das Besondere an der Datenbank Caché ist der Objektorientierten und SQL-basierten Zugriff auf Daten. Das gibt dem Programmierer die Freiheit zu entscheiden, je nachdem welcher Zugriff was am schnellsten zu Lösung führt.

Jedes anzuschließende System wird mit einem Adapter an Ensemble gekoppelt. Adapter sind Schnittstellenlösungen, die von Ensemble bereitgestellt werden und eine vereinfachte Sicht auf die häufig komplexen Schnittstellen darstellen. Zurzeit existieren über 250 Adapter, die die Kommunikation mit vielen Datenbanken und Anwendungen ermöglichen, z.B. SQL, HL7, TCP/IP und Webservice. Außerdem bietet Ensemble die Möglichkeit, Geschäftsprozesse mittels Business Processing Language (BPL) abzubilden. Mit Hilfe solcher Geschäftsprozess ist es möglich, die von den Adaptern bereitgestellten Daten zu aggregieren, zu transformieren und weiterzuverarbeiten [5].

Die Objektdatenbank mit ihrer proprietären Programmiersprache Caché ObjectScript wurde genutzt, um den Sorted Neighborhood Algorithmus zu implementieren. Zum Identifizieren von Duplikaten wurden die Distanzverfahren Levenshtein-, Damerau-Levenshtein- und Jaro-Winkler-Distanz implementiert.

3. Ergebnisse

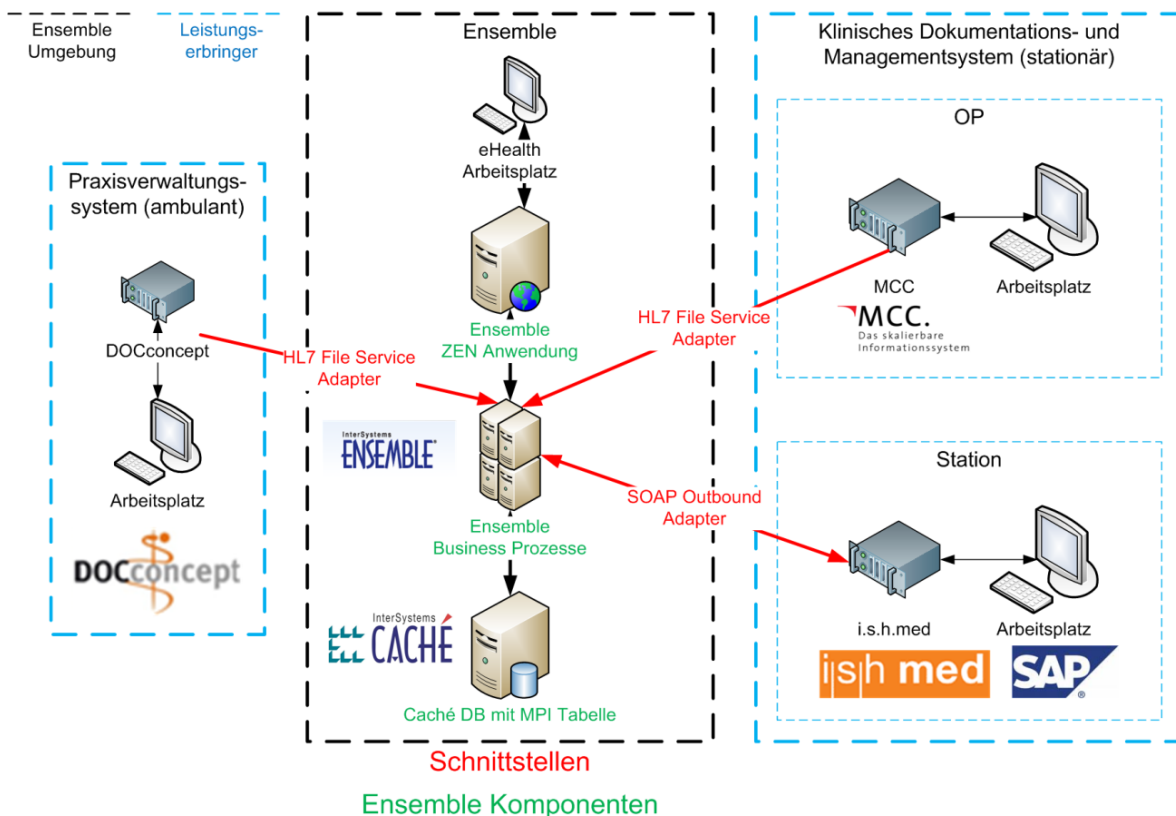


Abbildung 2: Architektur einer MPI-Umsetzung mit Ensemble

Über die Integrationsplattform Ensemble wurden die drei genannten Anwendungssysteme miteinander verbunden. Es kamen verschiedene Schnittstellen zum Einsatz. In der Übersicht auf *Abbildung 2* sind die einzelnen Schnittstellenumsetzungen für jedes Anwendungssystem und die Komponenten innerhalb von Ensemble dargestellt.

Die Anwendungssysteme DOCconcept und MCC wurden mittels HL7 basierter Schnittstellenadapter an Ensemble gebunden. Ensemble unterstützt als Integrationsplattform die Weiterleitung von HL7 Nachrichten anhand von Routing Regeln und die Transformation von HL7-Nachrichten mittels einer XML-basierten Data Transforming Language. Schnittstellenspezifische HL7-Nachrichten von angeschlossenen Anwendungssystemen konnten so in standardisierte Nachrichten umgewandelt werden, um Integrität und Weiterverarbeitung zu gewährleisten.

Die Transaktionen für bestimmte Anwendungsfälle, die bei der Umsetzung eines MPI eine Rolle spielen, wurde mit Hilfe des HL7-Profiles Patient Identifier Cross-referencing (PIX) der IHE umgesetzt. Der Prototyp unterstützt zurzeit die Transaktion Patient Identity Feed dieses Profils, und befüllt eine Tabelle der MPI Datenbank mit Patientenidentifikationsdaten. Dazu fängt Ensemble ADT-Nachrichten ab, die von den angeschlossenen Anwendungssystemen generiert wurden, und liest Felder aus den Nachrichten, speziell aus dem PID Segment, aus. Die ausgelesenen Identifikationsdaten für einen Patienten werden anschließend im MPI gespeichert. Es werden ausschließlich demographische Daten eines Patienten gespeichert, behandlungsbezogene Daten werden nicht in den MPI aufgenommen.

Neben den HL7-Schnittstellen der Systeme DOCconcept und MCC wurde die Webservice-Schnittstelle von IS-H/i.s.h.med angesprochen. Identifikationsdaten aus dem Patientenverwaltungssystem IS-H konnten über die Verwendung von Remote Procedure Calls (RPC) bereitgestellt werden. Ensemble stellt für den Webservice-basierten Nachrichtenaustausch auf Grundlagen des Simple Object Access Protokolls (SOAP) einen speziellen Adapter bereit, um Anfragen an ein externes System zu stellen und Antwortnachrichten zu empfangen. Die Antwortnachrichten können in ähnlicher Weise ausgewertet werden wie HL7-Nachrichten. Die Identifikationsdaten wurden ebenfalls in der MPI-Tabelle abgelegt. Zurzeit unterstützt der MPI die Transaktion Patient Identity Feed. Die ankommenden Patientendaten werden nach mehreren Attributen (Name, Vorname, Adresse, Quellsystemname...) gefiltert und es wird daraufhin ein Master Patient ID generiert. Da die ankommenden Daten aus den Quellsystemen Fehler (Tippfehler bei der Aufnahme) enthalten können, konnte auf keinen normaler Vergleich beim Einpflegen der Daten zurückgegriffen werden. Es wurde versucht, durch den Einsatz bestimmter Algorithmen, einen effektiveren Duplettenvergleich zu realisieren.

Die Patientendatensätze der MPI-Tabelle dienen als Testdaten, um die Funktionalität des Algorithmus zur Dublettenerkennung zu überprüfen. Durch die provozierte Erzeugung von Datensätzen mit Tippfehlern wurden die Effizienz und Performance der Implementierung getestet, um Aussagen über wahrscheinliche Duplikate und Laufzeiten treffen zu können. Durch Anpassung der Gewichtung einzelner Distanzmaßverfahren in Abhängigkeit nicht gefundener Duplikate kann die Implementierung nachträglich an spezifische Erfordernisse angepasst werden.

Im Zuge der Unterscheidung zwischen tatsächlichen Dubletten und Datensätzen die keine Dubletten voneinander darstellen, müssen Fehlerklassen eingeführt werden. Diese Fehlerklassen sind gleichzusetzen mit Schranken, welche Aussagen über den Grad einer Wahrscheinlichkeit treffen können. Vorliegende Implementierung unterstützt einen unteren Schwellwert von 85%, dabei wird der tatsächlich erreichte Score für eine Übereinstimmung gegen über dem vereinbarten Schwell-

wert abgeglichen. Fällt der beschriebene Score unter diesen Schwellwert kann davon ausgegangen werden, dass keine Übereinstimmung vorliegt. Die Vergleichenen Datensätze sind sich zu unähnlich. In der Praxis wird hier mit mehreren Schwellwerten gearbeitet.

4. Schlussfolgerung

Durch die Vielzahl an Schnittstellenumsetzungen und Nachrichtenschemata ist ein Transport von Patientendaten über Einrichtungsgrenzen ohne eine geeignete Middleware nicht möglich. Durch die Nutzung eines MPI ist die einrichtungübergreifende Patientenidentifikation möglich. Mit einem MPI und den in der Datenbanktabelle gesammelten Identifikationsdaten könnte unter Einsatz eines Dokumentenmanagementsystem, wie es im HL7 Profil Cross-Enterprise Document Sharing (XDS) der IHE beschrieben wird, eine vollständige Infrastruktur für den einrichtungübergreifenden Dokumentenaustausch geschaffen werden [4].

Die MPI-Clearingfunktion mit Hilfe des Sorted Neighborhood Algorithmus erfordert eine Balance zwischen Laufzeitverhalten und Erkennungsrate. Durch Vergrößerung des Sliding Window steigt zwar die Erkennungsrate, allerdings verschlechtert sich das Laufzeitverhalten exponentiell. Eine Vergrößerung des Sliding Window ermöglicht genauere Aussagen über in der Datenbanktabelle befindliche, wahrscheinlich ähnliche Duplikate. Durch Testen und gleichzeitiges Anpassen der Gewichtung muss eine befriedigende Erkennungsrate für die jeweils vorherrschende Anforderungen bei annehmbarer Performance ermittelt werden.

5. Literatur

[1] DAMERAU, F., A technique for computer detection and correction of spelling errors. Communications of the ACM.7 (Nr. 3), 1964.

[2] HEALTHCARE, SIEMENS: SOARIAN IC: Information zum Master Patient Index. PDF, 2007, http://www.medical.siemens.com/siemens/de_DE/gg_gs_FBAs/files/pdfs/Information_Technologies/SIC_ProductBrochure_EinlegerMPI_d_2007.pdf, Zitat vom 25. September 2009

[3] HÖFERT, F., Gesundheitmodernisierungsgesetz - Von Fall zu Fall - Pflege im Recht, 2009

[4] INTEGRATING THE HEALTHCARE ENTERPRISE® (IHE), IHE IT Infrastructure (ITI) Technical Framework, http://www.ihe.net/Technical_Framework/upload/IHE_ITI_TF_6-0_Vol1_FT_2009-08-10-2.pdf, 2009, Zitat vom: 31. August 2009.

[5] INTERSYSTEMS, GMBH, Ensemble – Die nächste Generation der Integrations Software. Webseite, <http://relaunch.intersystems.de/ensemble/ensemble-next-generation.html>, 2008, Zitat vom 15. September 2009.

[6] JARO, M.A., Probabilistic linkage of large public health data file. Statistics in Medicine 14 (5-7) 491–8. <http://www.census.gov/srd/papers/pdf/rrs2006-02.pdf>, 1995.

[7] LEVENSHTAIN, V.I., Binary codes capable of correcting deletions, insertions, and reversals., Soviet Physics Doklady, 1966.

[8] MONGE, A. und ELKAN, C., An Efficient Domain-Independent Algorithm for Detecting Approximately Duplicate Database Records. Proceedings of the Workshop on Research Issues on Data Mining and Knowledge Discovery, 1997.

[9] VHTG E.V., Konzept zur Patientenidentifikation. PDF, http://download.vhitg.de/Leitfaden_VHitG_PID_v09.pdf, 2006, Zitat vom 17. August 2009.

Corresponding Author

Anke Häber

University of Applied Sciences

Dr.-Friedrichs-Ring 2A, 08056 Zwickau

Email: anke.haerber@fh-zwickau.de