

## SynnX Knowledge Cockpit und SynnX Clinical Consult Application - passgenaue Wissenslösungen für Klinik und Forschung auf der Basis der Fingerprint-Textmining-Technologien

Herzog C<sup>1</sup>, Diwersy M<sup>1</sup>, Schmidt M<sup>1</sup>  
<sup>1</sup> SynnX Solutions GmbH, Köln, Deutschland  
christian.herzog@synnX.de

**Einleitung und Fragestellung** Medizinisch wissenschaftliches Fachwissen ist sowohl im klinischen Behandlungsprozess als auch im wissenschaftlichen Arbeitsprozess unentbehrlich. Allerdings haben Arzt und Wissenschaftler unterschiedliche Anforderungen an die Art des Fachwissens, an die Präsentation des Wissens, die Art und Weise, wie das Wissen in den Arbeitsprozess integriert werden muss und an Aktualität und Validität des Fachwissens. Während der Wissenschaftler große Publikationsmengen zu seinem Themengebiet überblicken, neueste Erkenntnisse und Trends frühzeitig entdecken, nach Möglichkeit mit intelligenten Methoden Wissen "zwischen den Zeilen" finden muss und vor allem assoziativ arbeitet, ist der klinisch tätige Arzt darauf angewiesen, im Arbeitsprozess ohne Zeitverzug und Zeitaufwand auf wenige ausgewählte behandlungsrelevante Daten zugreifen zu können. Diese müssen außerdem auf die speziellen Bedürfnisse des Patienten und die Arbeitssituation im Behandlungsprozess zugeschnitten sein. Beide Zielgruppen greifen auf den gleichen Wissenspool zu - aber mit deutlich unterschiedlichen Anforderungen, Herangehensweisen und Erwartungen.

SynnX hat mit dem SynnX Knowledge Cockpit für den Wissenschaftler und der SynnX Clinical Consult Application für den Arzt adäquate "Wissenslösungen" entwickelt, die die beschriebenen Anforderungen auf der technischen Basis der Collexis Fingerprint-Technologie umsetzen.

**Technische Basis** Technische Basis beider Anwendungen ist die Collexis Fingerprint-Engine, eine ontologie-basierte Indexierungs- und Retrieval-Lösung, die speziell für den biomedizinischen Bereich entwickelt wurde [1]. Auf dieser Basis werden unstrukturierte wissenschaftliche Texte, z.B. PubMed-Abstracts in sog. Fingerprints verwandelt, die ontologie-basiert identifizierten Fachbegriffe und Bewertungsrelationen für jeden Begriff enthalten. Bei der Indexierung und in den Anwendungen kommen mehrere Ontologien zum Einsatz, MeSH (Medical Subject Headings), UMLS (Unified Medical Language System) und ein speziell entwickelter Gen-Thesaurus [2].

Für die automatisierte Recherche von Fachpublikationen für den klinisch tätigen Arzt wurden die ICD-Codes in den MeSH-Thesaurus gemappt, um aus den Daten eines Patienten einen aussagefähigen "Patientenfingerprint" zu generieren, der eine automatisierte Recherche erlaubt. Dazu wurden die englischen Erläuterungstexte als Ausgangsbasis genommen, speziell definierte Stopwords entfernt und eine Indexierung mit der Collexis Core Engine und MeSH und UMLS als Thesaurus durchgeführt. In einem mehrstufigen Optimierungsprozess werden die Indexierungsergebnisse, die sich auf Thesaurus-Begriffen und Volltext-Begriffen zusammensetzen, regelbasiert in ICD-Fingerprints umgewandelt.

**SynnX Knowledge Cockpit - Schnelle Analyse großer Publikationsmengen** Mit dem SynnX Knowledge Cockpit wurde für den Wissenschaftler im Bereich Medizin / Life Science eine Lösung geschaffen, die ihm erlaubt, große Publikationsmengen zu einem definierten Thema schnell und effizient inhaltlich zu analysieren und das für ihn und seinen wesentliche herauszufiltern. Datenbasis eines Knowledge Cockpits sind wissenschaftliche Publikationen, z.B. als PubMed-Abstracts vorliegend, die eine thematische Vertikale definieren und dabei bis zu bis zu 100.000 Publikationen umfassen können, z.B. zum Thema Alzheimer. Diese Abstracts werden mit Collexis-Engine in Fingerprints umgewandelt und unter einer ontologiebasierten Navigation zugänglich gemacht. Gleichzeitig werden aus den Abstracts und den Meta-Informationen Expertenprofile generiert, die einen klaren Überblick erlauben, wer mit wem und zu welcher Zeit weltweit zum Thema des Cockpits gearbeitet hat. In der aus den Fingerprints generierten Navigation können nun schnell relevante Publikationen identifiziert werden, z.B. die Publikationen die neben dem zentralen Thema eines Cockpits auch eine weitere Erkrankung und Medikament enthalten.

Im Bereich Knowledge Discovery wird für den Wissenschaftler aufbereitet, welche Begriffe erstmalig in einer Veröffentlichung mit dem Cockpit-Thema genannt werden und welche Fachbegriffe voraussichtlich mit dem Cockpit-Thema in Zukunft gemeinsame Relevanz entwickeln werden. Die Basis für die Analysen bildet ein Vektorraum aus 5 Millionen Konzeptvektoren abgeleitet aus den PubMed-Abstracts seit 1996. Dieser Dokumentenraum wird durch Analyse der Konzeptassoziationen (gemeinsames Auftreten in mindestens einer Publikation) in einen Assoziationsraum gewandelt. Dieser Vektorraum, bestehend aus ca. 35.000 Dimensionen (den benutzten MeSH-Konzepten) und 60 Millionen (direkten) Assoziationen, kann nun dafür genutzt werden, um mittels Ähnlichkeitsanalyse neue Assoziationen zu finden, die in dieser Kombination so noch nicht in der Literatur gebildet wurden. Durch Interpolation der Ähnlichkeitsfaktoren über einen Zeitstrahl von 10 Jahren können feinste Bewegungen in diesem Raum analysiert werden und Wahrscheinlichkeiten für entstehende Assoziationen berechnet werden. Das Ergebnis sind gewichtete Vorschlagslisten von potentiellen (also indirekten) Assoziationen, die es einem Experten erlauben, z.B. Substanzen passend zu einem bestimmten Symptom zu entdecken oder eventuelle gewünschte oder unerwünschte Nebenwirkungen einer Substanz früher in seine Überlegungen einzubeziehen als es ohne diese Analyse möglich wäre.

**SynnX Clinical Consult Application - passgenaues Wissen im klinischen Kontext** Mit der SynnX Clinical Consult Application werden dem Arzt für den jeweiligen Patienten relevante Fachpublikationen mit klinischer Relevanz und direkt im Arbeitsprozess angezeigt[3]. Möglich wird dies durch die direkte Anbindung an das Krankenhausinformationssystem (KIS) und die automatisierte Generierung eines "Patientenfingerprints" auf kodierten Informationen wie ICD-Codes, Altersgruppe und Geschlecht des Patienten. Dieser generierte Patientenfingerprint wird mittels der Collexis Core Engine mit den vorliegenden Fingerprints von wissenschaftlichen Publikationen verglichen. Die Veröffentlichungen mit einer definierten Ähnlichkeit werden dem Arzt im KIS oder im Intranet angezeigt. Dabei werden über Filter, die klinisch relevante Publikationen erkennen lassen, und über den Publikationstyp der Arzt mit behandlungsrelevanten Dokumenten versorgt (z.B. Cochrane Reviews als leitlinienähnliche Dokumente etc.) oder Case Reports auf PubMed. Der Behandler kann die Sicht durch die Ergänzung von angebotenen Schlüsselbegriffen verfeinern und anpassen. Mit der SynnX Clinical Consult Application, die in verschiedene KIS-Systeme integriert werden, werden dem Arzt im Arbeitsprozess ausgewählte wissenschaftliche Inhalte angeboten. Hauptaugenmerk liegt darauf, den Arzt im gedrängten klinischen Alltag nicht mit Informationen zu überfluten, sondern die Einbindung von Informationsangeboten in den Workflow den Arzt zur Auseinandersetzung mit Innovationen und Fachveröffentlichungen anzuregen

**Zusammenfassung** Mit den beiden beschriebenen Lösungen werden auf der Basis der Collexis Fingerprint Engine angepasste Wissenslösungen für den Wissenschaftler und den behandelnden Arzt umgesetzt, die die stark divergierenden Informationen entsprechend der Anforderungen präsentieren: für den Wissenschaftler wird über große Publikationsmengen eine strategischer Transparenz geschaffen, gleichzeitig wird der Wissenschaftler mit Knowledge Discovery-Auswertungen unterstützt. Im klinischen Bereich werden die Informationen auf der Basis von Patientenmerkmalen automatisiert recherchiert und der Arzt proaktiv versorgt.

### Literatur

- [1] van Mulligen EM, Diwersy M, Schmidt M, Buurman H, Mons B. Facilitating networks of information. Proc AMIA Symp. 2000;:868-72
- [2] Schijvenaars BJ, Mons B, Weeber M, Schuemie MJ, van Mulligen EM, Wain HM, Kors JA Thesaurus-based disambiguation of gene symbols. BMC Bioinformatics 2005 Jun 16, 6:149
- [3] Christian Herzog, Gianpiero Liuzzi, Mario Diwersy. SynnX solutions: practical knowledge management in a medical environment. Bremen: ACM Press, CIKM 2005: 556-559