

d-fine



Post-Market Surveillance: Klinische Studien effizienter überwachen

Post-Market Surveillance: Klinische Studien effizienter überwachen

Die Post-Market Surveillance (PMS) stellt Medizinproduktehersteller vor die Herausforderung, ihre Produkte betreffende Informationen in zahlreichen öffentlichen Datenquellen finden und zusammentragen zu müssen. Das Johner Institut hat sich zum Ziel gesetzt, eine moderne „One-Stop-Shop“-Lösung zu entwickeln, die mittels maschinellem Lernen und Automatisierung Mitarbeiter optimal bei der Recherche unterstützt und so deren Produktivität nachhaltig steigert. d-fine hat das Johner Institut hierbei durch die Entwicklung einer intelligenten Suchlösung für klinische Studien unterstützt: Dank gezielter Symbiose von maschineller und menschlicher Intelligenz werden relevante Studien gezielter gefunden und der Anwender automatisch über neue Studienergebnisse informiert.

Post-Market Surveillance: Ein häufig manueller Rechercheprozess mit strengen Anforderungen an Dokumentation und Qualität

Die Post-Market Surveillance ist im Kern eine strukturierte Suche nach Informationen zur Produktsicherheit. Diese Informationen müssen dabei aus verschiedenen Quellen, wie beispielsweise Verzeichnissen klinischer Studien, Mitteilungen von Herstellern und wissenschaftlichen Publikationen, zusammengetragen werden, eine zentrale Anlaufstelle gibt es nicht. Dabei werden an die Recherche hohe regulatorische Anforderungen hinsichtlich Vollständigkeit und Qualität gestellt und in Form von Standard Operating Procedures (SOPs) konkrete Arbeitsanweisungen formuliert. Die Arbeitsabläufe sind typischerweise manuell und gleichzeitig gerade hinsichtlich der Dokumentation standardisiert, wodurch sich ein großes Potential zur Produktivitätssteigerung durch eine optimierte Software-Unterstützung bietet. Um die Anforderung der Nutzer an eine moderne Lösung besser zu verstehen, werfen wir zunächst einen Blick auf die wohl erfolgreichste Such-Lösung aus dem Consumer Bereich: Google.

Mitarbeiter erwarten von Business Software die zu Hause gewohnte Leistungsfähigkeit und Nutzerfreundlichkeit von Google und Co.

Allseits präsenste Suchdienste wie beispielsweise Google verdeutlichen wie eine zielgerichtet gestaltete User Experience einen Nutzer unterstützten kann und repräsentieren zugleich das, was ein Mitarbeiter heute von einer digitalen Lösung erwartet. In der funktionellen Zielsetzung gibt es selbstverständlich signifikante Unterschiede: Während es ein Web-Suchdienst dem Nutzer ermöglicht, effizient einzelne, für ihn besonders interessante Inhalte, zu finden, muss eine PMS Recherche-Lösung eine umfassende und lückenlose Recherche in ausgewählten Quellen ermöglichen und den spezifischen Arbeitsprozess gezielt unterstützen. Wir können jedoch die vier zentralen Aspekte identifizieren, die adaptiert werden sollten, da sie den Nutzer schnell zu seinem gewünschten Suchergebnis führen:

- **„One-Stop-Shop“:** Ein einzelner Service bietet den Zugang zum gesamten Informationsangebot.
- **Ergebnis-Sortierung nach Relevanz:** Die Suchmaschine „erkennt“ welche Treffer für den Nutzer am interessantesten sind und zeigt diese zuerst. Ermöglicht wird diese maschinelle Intelligenz durch Machine Learning Methoden und aggregierte historische Daten.
- **Kontextualisierung:** Für jeden Treffer wird dem Nutzer der textliche Kontext, in dem seine Suchbegriffe gefunden wurden, dargestellt. Er kann entscheiden, ob ihn ein Treffer interessiert, ohne den vollständigen Inhalt durchforsten zu müssen.
- **Unterstützung:** Bei der Eingabe von Suchbegriffen werden Schreibfehler korrigiert und Synonyme verwendet, wobei der Nutzer jedoch stets korrigierend eingreifen kann.

Legt man diese elementaren Nutzeranforderungen zu Grunde und erweitert diese um PMS-spezifische Business-Anforderungen insbesondere an Vollständigkeit, Auditierbarkeit und Validierung, so gibt es auf dem Markt bisher keine Lösung, die alle Aspekte zufriedenstellend vereint.

Angebotene Business-Lösungen fallen hinter die berechtigten Erwartungen der Anwender zurück. Daher haben wir eine eigene Lösung entwickelt.

Prof. Dr. Christian Johner und das Johner Institut haben sich daher als Vorreiter im Medizintechnikmarkt mit dem Einsatz von maschineller Intelligenz und gezieltem User-Experience-Design (UX) zum Ziel gesetzt, eine moderne PMS Lösung zu entwickeln. Als Pilot für den Einsatz maschineller Intelligenz wurde d-fine mit der Entwicklung eines Suchservice für die Daten innerhalb des Registers für klinische Studien ClinicalTrials.gov (CT.gov) beauftragt. Dieses von der US Library of Medicine betriebene Register enthält Informationen über mehr als 330000 klinische Studien und ist ein wichtiger Bestandteil jeder PMS-Recherche, da die FDA eine Veröffentlichung neuer klinischer Studien in diesem Register vorschreibt. Die CT.gov-Website stellt für den Zugriff auf die Daten des Registers einen Suchdienst bereit¹. Der Fachanwender steht jedoch regelmäßig vor dem Problem, dass einerseits zu viele nicht-relevante Treffer gefunden und andererseits bekanntermaßen relevante Studien gar nicht gefunden werden. Für die Bewertung der Leistungsfähigkeit unserer neuen Lösung dient dieser Suchdienst als Benchmark, wobei wir das Potential zur Verbesserung an drei zentralen Stellen heben:

1 Präziserer Suchalgorithmus: mehr relevante Studien, weniger irrelevante Studien

2 Optimale Mensch-Maschine-Interaktion: effizienteres Arbeiten durch erklärbare maschinelle Intelligenz, aktive Kontrolle durch den Nutzer und Automatisierung regelmäßiger Aufgaben

3 Zukunftsfähige Technologie: moderne Toolchain und Basis für „One-Stop-Shop“-Lösung durch flexible Aggregation von verschiedenen Quellen

1. Unsere Lösung liefert fast doppelt so häufig relevante Treffer unter den Top 5 wie eine vergleichbare Webseitensuche

Anhand von 30 klinischen Bewertungen bzw. repräsentativen Suchanfragen wurde der Suchprozess unter Verwendung der neuen Lösung mit einer äquivalenten Suche auf der CT.gov-Website verglichen (Abbildung 1) und es ergaben sich überzeugende Ergebnisse. Insgesamt konnten 35% mehr relevante Treffer gefunden werden, wobei in 3 der 30 Bewertungen relevante Studien gefunden wurden, in denen bei der Vergleichssuche gar nichts gefunden wurde. Nicht-relevante Treffer wurden um 75% reduziert und dem Anwender damit erheblicher Zeitaufwand erspart. Betrachtet man die Top 5 Treffer einer Suche, welchen erfahrungsgemäß die größte Nutzeraufmerksamkeit zukommt, werden mit unserer Lösung fast doppelt so viele relevante Treffer gefunden.

Erreicht wurden diese Verbesserungen durch drei zentrale Komponenten: Die Volltexte aller Studiendaten lassen sich durch moderne Indizierungstechniken ohne Performanceverlust hinsichtlich Reaktionszeiten durchsuchen; Dabei werden Synonyme aus der Ontologie UMLS nicht nur automatisch einbezogen, sondern auch intelligent hinsichtlich ihrer Sinnhaftigkeit bewertet; Schlussendlich kommt ein leistungsstarker Sortier-Algorithmus zum Einsatz, der die Ergebnisse hinsichtlich ihrer Relevanz für den Nutzer sortiert. Die hierfür eingesetzte Methodik wird im nächsten Abschnitt beschrieben.

¹ <https://clinicaltrials.gov/ctz/search/advanced>

Vergleich relevanter Treffer – Webseitensuche

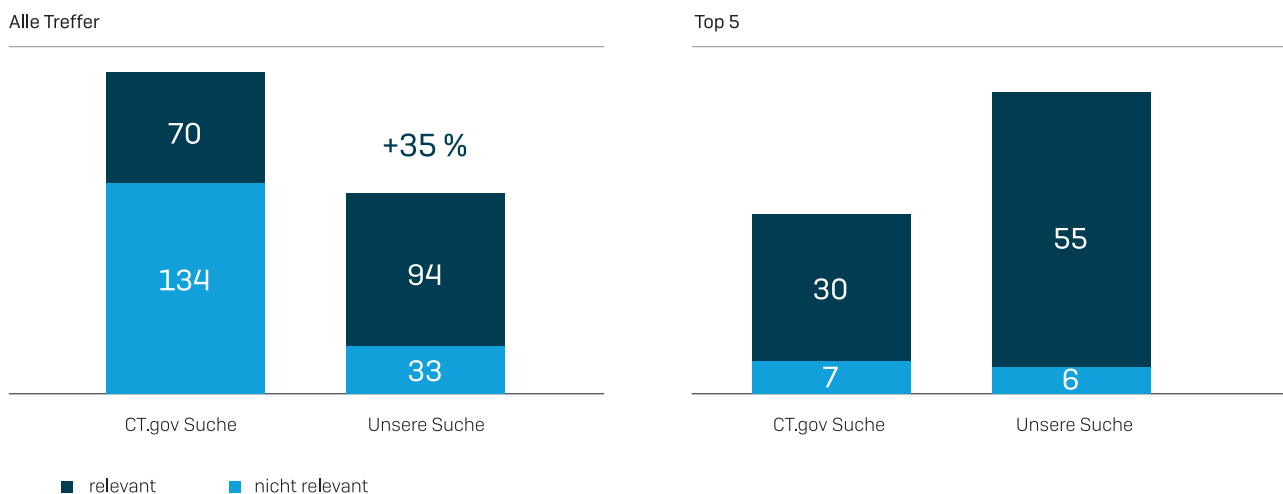


Abb. 1: Anzahl (nicht-) relevanter Treffer über die Webseitensuche (CT.gov Suche) und unsere Lösung aus 30 verschiedenen Suchanfragen. Betrachtet man jeweils die obersten 5 Treffer der sortierten Ergebnislisten (höchste Nutzeraufmerksamkeit), bietet unsere Lösung fast doppelt so viele relevante Treffer.

2. Unsere Lösung kombiniert leistungsstarke ML-Methoden mit größtmöglicher Verständlichkeit, um eine symbiotische Mensch-Maschine-Interaktion zu erreichen

Der Nutzer möchte die für ihn relevantesten Treffer zuerst angezeigt bekommen. Wir haben dafür ein intelligentes und zugleich für den Nutzer nachvollziehbares ML-Modell für die Relevanzsortierung entwickelt. Basierend auf mehr als 1000 klassifizierten Trainings-Datensätzen wurden die zentralen Einflussfaktoren auf die Relevanz gemessen und nachfolgend eine von Fachanwendern validierte Kategorisierung in 4 Stufen eingeführt (prime – high – medium – low).

Der Graph in Abbildung 2 zeigt einen Vergleich zwischen der automatischen Zuordnung in diese 4 Kategorien und einer Bewertung des Fachanwenders („relevant“ / „nicht relevant“), ob diese Einschätzung korrekt war. Die erfolgreiche Separation zwischen relevanten und irrelevanten Treffern lässt sich deutlich erkennen: Die obersten zwei der vier Kategorien weisen über 80% relevante Treffer auf. Die Bedeutung der Kategorien wird dem Anwender anhand einem der zentralen Einflussfaktoren im Modell erklärt: des Vorkommens der Suchbegriffe in bestimmten Datenfeldern. Abbildung 2 zeigt die Zuordnung, in welchem Kontext Begriffe gefunden wurden für jede Kategorie. Ein weiterer, zentraler Mehrwert dieses Bewertungsmodells ist seine Vergleichbarkeit. Da ein absolutes Ranking erstellt wird, können nicht nur die Ergebnisse einer Suchanfrage sortiert werden, sondern auch ver-

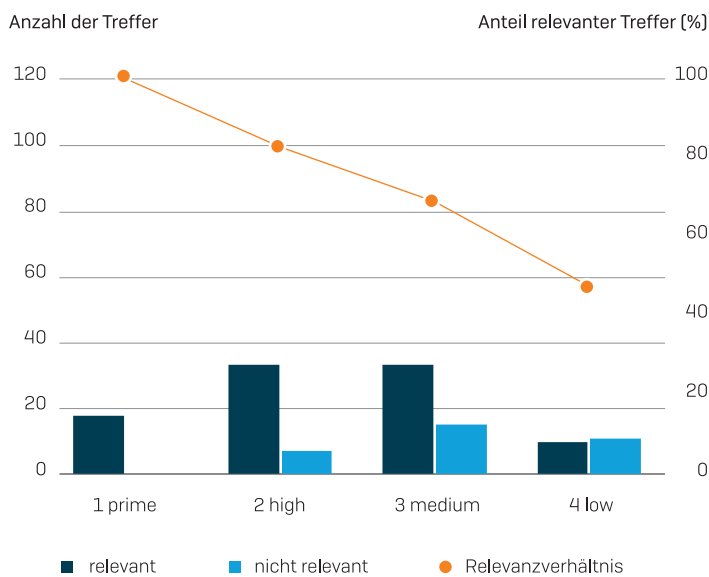
schiedene Suchanfragen miteinander verglichen werden. Beispielsweise kann ein Anwender erkennen, dass er seine Suchanfrage vermutlich treffender formulieren kann, wenn seine Recherche nach einem viel genutzten, lange am Markt existierenden Produkt ausschließlich Treffer der Kategorie 4 liefert.²

Die Verständlichkeit der maschinellen Entscheidungen hat direkten Einfluss auf die Anwenderfreundlichkeit, wobei die Kontrollhoheit des Nutzers als ein Eckpunkt eines guten UX-Designs stets Beachtung findet. Dank der Kontextualisierung der gefundenen Begriffe kann ein Anwender Ergebnislisten schneller abarbeiten und das System ebenfalls bei Fehleinschätzungen korrigieren. Die automatisierte Verwendung von Synonymen wird jederzeit deutlich dargestellt und kann vom Nutzer angepasst werden. Darüber hinaus wird dem Anwender die regelmäßige Recherche und Dokumentation seiner Arbeit erheblich vereinfacht, indem Suchanfragen gespeichert und automatisiert werden können. Sobald ein neuer Treffer gefunden wird, wird der Nutzer benachrichtigt. Alle Vorgänge werden dabei nachverfolgbar und auditsicher abgespeichert.

Die enge Verzahnung zwischen Software und Anwender erlaubt das strukturierte Sammeln von Feedback ohne Mehraufwand für den Anwender und mit der Zeit können so wachsende Trainings-Datensätze aufgebaut werden. Ein stetiger Ausbau der überwacht lernenden ML-Methodik ist damit möglich.

² CT.gov bietet lediglich eine relative Relevanzsortierung, ein Vergleich kann hier daher nicht durchgeführt werden.

Trefferwahrscheinlichkeit nach Relevanzkategorie



Beschreibung der Kategorie

1 prime	Vollständiger Titel oder Kurztitel
2 high	Zusammenfassung, Feld „conditions“ oder Feld „interventions“
3 medium	Detailbeschreibung, Design- oder Ergebnisgruppen und Schlüsselworte
4 low	Anderer Inhalt und lexikalische Varianten (z.B. keine Bindestriche)

Abb. 2: links: Die Gegenüberstellung von maschineller Einteilung der obigen Suchergebnisse in vier Kategorien (1-4) und Bewertung des Fachanwenders (relevant / nicht relevant) demonstriert die hohe Trennschärfe des maschinellen Bewertungsalgorithmus. Das Verhältnis der jeweiligen Trefferanzahl ergibt den dargestellten Anteil relevanter Treffer pro Kategorie. rechts: Erklärung der Zuordnung zu den Relevanzkategorien. Zentraler Faktor ist das Vorkommen des gesuchten Begriffes in bestimmten Datenfeldern.

3. Unsere technische Umsetzung basiert auf einer modernen Tool-Chain und ist auf kontinuierliche Veränderung ausgelegt

Die technische Umsetzung unserer Lösung ist in Abbildung 3 skizziert: Eine Mikroservice-Architektur, in Docker-Containern bereitgestellt, bietet ein leicht zu wartendes und erweiterbares System und stützt sich auf etablierte und moderne Open-Source-Tools. Alle Funktionen des Such-Services werden über eine Web-API bereitgestellt und sind vollständig von der Nutzeroberfläche entkoppelt. Somit lässt sich die Lösung sowohl in die PMS-Plattform des Johner Instituts als auch bequem in andere Anwendungen integrieren. Der Bezug der Synonyme aus gängigen Ontologien wie UMLS ist als eigener API-Service realisiert und bietet neben der ver-

einfachten Anbindung der UMLS-API eine Bewertung und Kategorisierung der bezogenen Synonyme. Diese Realisierung ermöglicht, dass neben dem Such-Service auch andere Anwendungen innerhalb der Plattform von Synonymen profitieren können.

Die technische Architektur wurde vor dem Hintergrund gewählt, dass eine stetige Entwicklung des Systems stattfinden wird. So können weitere Quellen modular und flexibel integriert werden, wie beispielsweise Publikationen aus MedLine über PubMed und Veröffentlichungen der FDA. Weiterhin ist der Bewertungskern erweiter- und austauschbar, sodass mit wachsender Datenbasis weitere ML-Methoden nahtlos integriert werden können und deren Steuerung über ein ML-Lebenszyklus-Tool verwaltet und automatisiert werden kann.

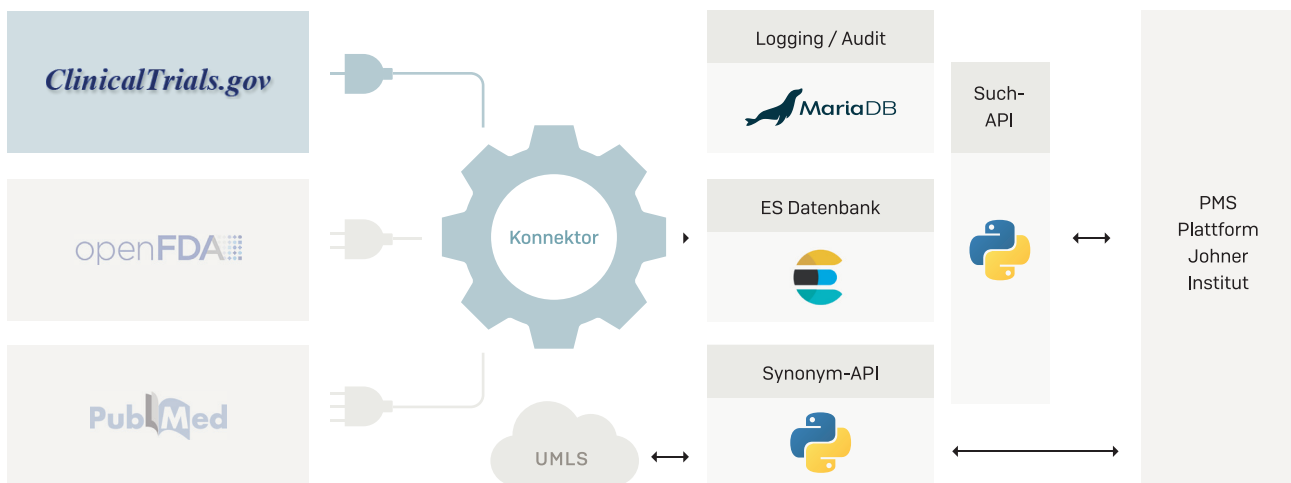


Abb. 3: Skizze der Lösungsarchitektur: Quelldaten werden in einer ElasticSearch (ES) Datenbank aggregiert und der Such-Service und Synonym-Service über eine Web-API bereitgestellt. Eine relationale Datenbank dient darüber hinaus zur Aufnahme von Audit-Trails. Zunächst wurde ClinicalTrials.gov als Datenquelle angebunden, weitere Quellen (z.B. openFDA oder PubMed) können zukünftig flexibel hinzugefügt werden.

Zusammenfassung und Ausblick

Um den Prozess der Post-Market Surveillance effizient zu unterstützen, haben wir ein flexibles, erweiterbares System entwickelt, welches eine signifikant verbesserte Performanz bietet und von Grund auf mit einem Fokus auf ein gutes UX-Design gestaltet wurde. Der gewählte konzeptionelle Ansatz stiftet von Anfang an Mehrwert,

statt über eine lange Investitionsstrecke auf das Zusammentragen eines sehr großen Datensatzes angewiesen zu sein. Die modulare Architektur erlaubt eine flexible Integration weiterer Datenquellen, sowie eine einfache Integration in Anwendungen, während das strukturierte Anwenderfeedback einen stetigen Ausbau der überwacht lernenden ML-Methoden und eine stetige Verbesserung des Suchalgorithmus ermöglicht.

Kontakt

Sie stehen vor ähnlichen Herausforderungen und möchten sich mit den d-fine Data Analytics Experten austauschen?

Sprechen Sie uns an!

pharma_and_healthcare@d-fine.com

Autoren

DR. MARTIN SCHNACK

Consultant, d-fine GmbH, Frankfurt
martin.schnack@d-fine.de

DR. ROBERT GÖRKE

Manager, d-fine GmbH, Frankfurt
robert.goerke@d-fine.de

DR. BODO HUCKESTEIN

Partner Pharma & Healthcare
d-fine GmbH, Frankfurt
bodo.huckestein@d-fine.de

Wir bedanken uns herzlich bei **Dr. Daniel Lohner** für fruchtbare Diskussionen und seine Beiträge bei der Entstehung dieses Whitepapers.

Berlin

d-fine GmbH
Friedrichstraße 68
10117 Berlin
Deutschland
berlin@d-fine.de

Düsseldorf

d-fine GmbH
Dreischeibenhaus 1
40211 Düsseldorf
Deutschland
duesseldorf@d-fine.de

Frankfurt

d-fine GmbH
An der Hauptwache 7
60313 Frankfurt
Deutschland
frankfurt@d-fine.de

München

d-fine GmbH
Bavariafilmplatz 8
82031 Grünwald
Deutschland
muenchen@d-fine.de

London

d-fine Ltd
6-7 Queen Street
London, EC4N 1SP
United Kingdom
london@d-fine.co.uk

Wien

d-fine Austria GmbH
Riemergasse 14 Top 12
1010 Wien
Österreich
wien@d-fine.at

Zürich

d-fine AG
Brandschenkestrasse 150
8002 Zürich
Schweiz
zuerich@d-fine.ch