

Ein Technologie-Scouting-System für Recherche und Vernetzung im Schienenverkehrssektor

Die Breite der für den Schienenverkehrssektor relevanten Technologien ist groß, dynamisch und übersteigt die Kapazitäten für manuelle Recherche im Arbeitsalltag. Intelligente Textanalyse-Methoden können diese Rechercheaufgaben teilweise automatisiert übernehmen. Kombiniert mit menschlicher Expertise bei der Qualitätssicherung kann so ein leistungsfähiges Technologie-Scouting-System entstehen.



Die Menge an Informationen zu Technologieinnovationen und deren kontinuierliches Wachstum stellen das Wissens- und Innovationsmanagement im Schienenverkehrssektor vor immer größere Herausforderungen. Das Innovationstempo in der Branche steigt, das Spektrum relevanter Technologien wird breiter – von nachhaltigen Antriebsformen [1, 2] über Materialentwicklung [3] und neuen Technologien für Energieeffizienz [4] bis zum automatisierten Bahnbetrieb [5] – und Technologien aus anderen Industrien werden zunehmend auch für den Schienenverkehr relevant. Für verschiedene Akteure ist es deshalb wichtig, einen Überblick über Technologien und deren Entwicklung zu behalten. Unternehmen benötigen bspw. Einschätzungen zu Reifegrad und Einsatzgebieten von Technologien, um ihre Beschaffung auszurichten, Kooperationen einzugehen oder für die Produktentwicklung. Forschungsorganisationen und Fördergeber müssen Technologien und deren potenzielle Anwendungen einordnen können, um zielgerichtet Programme entwickeln zu können.

Die systematische Suche, Analyse und Einordnung von Technologien wird als Technologie-Scouting (auch „Horizon Scanning“ oder „Technology Intelligence“) bezeichnet. Ansätze und Anwendungen für das Technologie-Scouting existieren bereits seit längerem und in unterschiedlichen Branchen. Im Zuge der Digitalisierung der Informationen zu Technologien bietet

sich die Möglichkeit, bestimmte Teilaufgaben des Technologie-Scoutings wie etwa das Recherchieren, Filtern und Zusammenführen von Informationen mit Hilfe künstlicher Intelligenz (KI) maschinell durchzuführen und so das Technologie-Scouting in Teilen zu automatisieren. Menschliche Expertise kann gezielt z. B. bei der Überprüfung und Ergänzung von Wissen zum Einsatz kommen. So lassen sich die jeweiligen Stärken menschlicher und maschineller Methoden kombinieren.

Im Schienenverkehrssektor kann ein solches teilautomatisiertes Technologie-Scouting-System (TSS) verschiedene Akteure dabei unterstützen, den Überblick über Technologieentwicklungen zu behalten. Dabei stellen sich jedoch noch viele Fragen, die von den anzuschließenden In-



Frederick Blumenthal
d-fine GmbH,
Frederick.Blumenthal@d-fine.de



Dr. rer. nat. Thomas Buder
Deutsches Zentrum für
Schienenverkehrsforschung
beim Eisenbahn-Bundesamt,
BuderT@dzsf.bund.de



Dr. rer. nat. Lisa Löbling
d-fine GmbH,
Lisa.Loebing@d-fine.de



Dr. rer. nat. Thorsten Sickenberger
d-fine GmbH,
Thorsten.Sickenberger@d-fine.de



Dr. rer. nat. Rustam Tagiew
Deutsches Zentrum für
Schienenverkehrsforschung
beim Eisenbahn-Bundesamt,
TagiewR@dzsf.bund.de



Dr. rer. pol. Roman Tilly
Deutsches Zentrum für
Schienenverkehrsforschung
beim Eisenbahn-Bundesamt,
TillyR@dzsf.bund.de



Dr. phil. nat. Phillip Yao-Lakaschus
d-fine GmbH,
Phillip.Yao-Lakaschus@d-fine.de

formationsquellen über die Darstellungsformen und die einzusetzenden KI-Verfahren im System bis zur genauen Verteilung der Aufgaben zwischen Mensch und Maschine und der Gestaltung eines langfristig geeigneten Betreibermodells reichen. In einer Vorstudie haben das DZSF und d-fine konzeptionell und prototypisch untersucht, ob und wie ein teilautomatisiertes TSS für den Schienenverkehrssektor aufgebaut und betrieben werden könnte.

1. Teilautomatisiertes Technologie-Scouting – was gehört dazu?

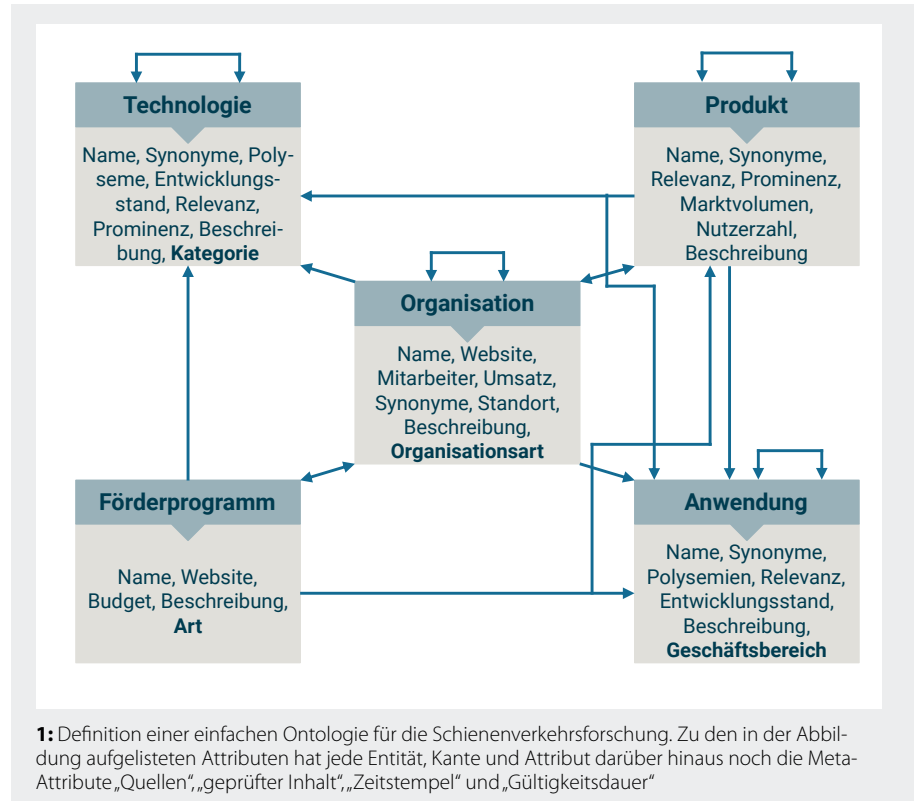
Bei der Konzeption eines TSS sind vier Aspekte essenziell und bestimmen wesentlich die Einsatzmöglichkeiten:

1. die Art und Struktur des Wissens,
2. die genutzten Datenquellen,
3. die maschinellen Methoden für die Wisensextraktion und
4. die Darstellungen und Funktionalitäten für die Beteiligung der Nutzenden.

1.1. Eine Ontologie für den Schienenverkehr

Die Struktur des Wissensnetzwerks wird durch eine Ontologie beschrieben. Sie definiert, welche Entitätstypen (Klassen) mit welchen Eigenschaften und welche Beziehungen zwischen den Klassen im System abgebildet werden können. Für die Vorstudie wurde eine Ontologie (s. Bild 1) entworfen, die die Klassen „Technologie“ (z.B. Brennstoffzelle), „Organisation“ (z.B. Eisenbahn-Bundesamt), „Anwendung“ (z.B. Fahrzeugantrieb), „Produkt“ und „Förderprogramm“ (z.B. HORIZON EUROPE) umfasst. Im TSS können dann zu einzelnen Instanzen dieser Klassen Informationen und Beziehungen untereinander (z.B. „Organisation A forscht an Technologie B“, „Förderprogramm C fördert Anwendung D“) gespeichert werden.

Zu den Technologien können Eigenschaften automatisiert aus den Informationsquellen herausgelesen werden. Dazu zählen Synonyme, falls Technologien unter verschiedenen Begriffen bekannt sind (z.B. „DAC“ und „digitale automatische Kupplung“). Ebenso können geographische Informationen (z.B. Länder, Städte) zugeordnet werden, um so die Verbreitung von Technologien zu verfolgen. Komplexere Charakteristika wie „Relevanz“ und „Entwicklungsstand“ erfordern aufwendigere Methoden, die textuelle Zusammenhänge erkennen können.



1.2. Die Quellen des Wissens – nicht nur Fachpublikationen

Für die automatisierte Extraktion von Informationen bieten sich verschiedene Arten von Quellen an. Wissenschaftliche Publikationen aus der Schienenverkehrsforschung unterliegen einer Qualitätssicherung, sodass Informationen belegt und belastbar sind. Um eine hohe Detailtiefe zu erreichen, sind gerade solche Veröffentlichungen mit starkem Sektorfokus relevant. Andere Publikationsarten wie Social Media sind zwar wenig fachspezifisch und nicht in gleichem Maße qualitätsgesichert, zeichnen sich aber durch ihre Aktualität aus. Patente und Datenbanken zu Förderprogrammen bieten eine gute Ergänzung, um eine breite Abdeckung der Ontologie zu erreichen. In der Vorstudie wurden über 70 Einzelquellen für ihren Einsatz in einem TSS untersucht.

Neben den inhaltlichen Aspekten, dem Umfang und der Veröffentlichungsfrequenz sind auch technische Aspekte für die Auswahl von Quellen relevant. Von Vorteil ist, wenn Informationen mit geringem Aufwand maschinell ausgelesen werden können. Besonders Quellen, die Inhalte und Meta-Informationen in maschinenlesbarem Format (z.B. JSON, XML) über Programmierschnittstellen bereitstellen, eignen

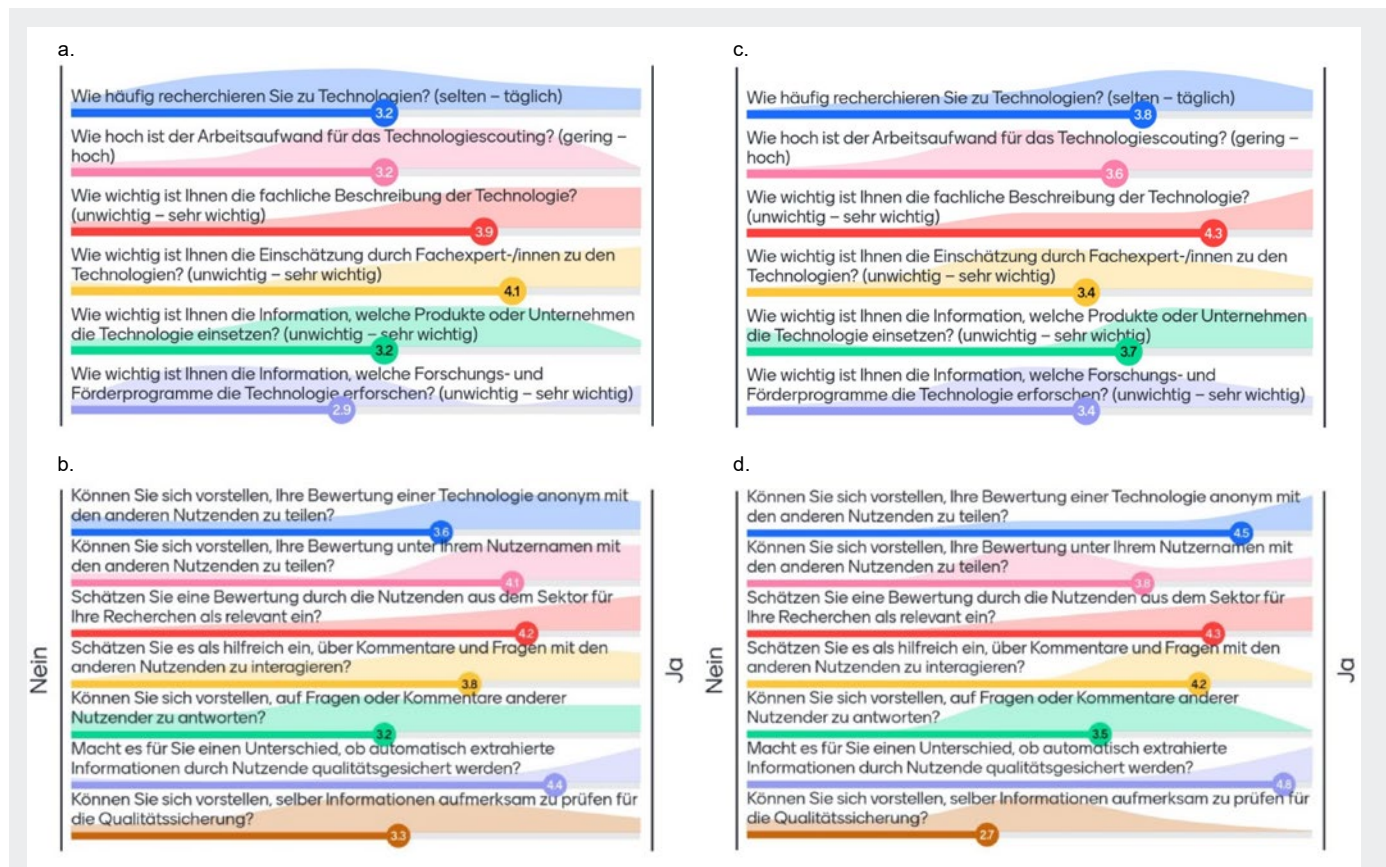
sich für eine automatisierte Anbindung. Technisch anspruchsvoller ist das Auslesen von Texten aus zwar digitalen, aber nicht strukturierten Dokumenten (z. B. PDF).

Die Nutzung von Quellen für die automatisierte Wisensextraktion mittels Text and Data Mining¹⁾ (TDM) unterliegt urheberrechtlichen Vorgaben. Die gesetzlichen Schranken sind dabei breiter gefasst, wenn TDM zum Zwecke wissenschaftlicher Forschung erfolgt. Die engerere, allgemeine TDM-Schranke regelt die möglichen Nutzungshandlungen auch für TDM zu kommerziellen Zwecken. Um eine Einschränkung des Anwendungsbereichs auf die wissenschaftliche Forschung zu vermeiden, sollte man sich für ein TSS für den Schienenverkehrssektor lediglich der allgemeinen TDM-Schranke bedienen.

1.3. Natural Language Processing – Wisensextraktion mithilfe von KI

Um aus den unstrukturierten Quelltexten automatisiert Wissen zu gewinnen, sind geeignete Methoden des Natural Language Processing (NLP) nötig. Dabei sind eine Vielzahl von Unteraufgaben zu lösen, z.B. das Erkennen und Extrahieren der Entität

1) Automatisierte Analyse von Inhalten, um Informationen und Korrelationen zu extrahieren



2: Antwortverteilung hinsichtlich Relevanz von Technologie-Scouting und Informationen sowie zu verschiedenen Beteiligungsmöglichkeiten. Links: Verteilung der Antworten im ersten Workshoptermin (a. n=19, b. n=14), rechts: im zweiten Workshoptermin (c. n=9, d. n=6)

ten der Ontologie wie Technologien und Organisationen (Named-Entity-Recognition), die Ableitung der Beziehungen zwischen Entitäten (Relation Extraktion) oder eine Kategorisierung von Technologien in z.B. „Rollmaterial“, „Infrastruktur“ und „Digitalisierung“. Für gewisse Aufgaben können bereits einfache, regelbasierte Modelle geeignet sein, während für komplexere Aufgaben Modelle aus dem KI-Teilbereich des maschinellen Lernens notwendig sind. In der Vorstudie wurde das Spektrum an Methoden zur Bewältigung der verschiedenen Teilaufgaben aufgearbeitet und deren Vor- und Nachteile beleuchtet.

Eine sehr hohe Genauigkeit kann durch die aufgaben- und domänenspezifische Fein Anpassung (Fine-Tuning) vortrainierter Sprachmodelle mit manuell annotierten Daten erreicht werden. Sehr große Sprachmodelle können bereits ohne aufgabenspezifisches Training (und somit ohne annotierte Daten) für viele Aufgaben eine angemessene Genauigkeit erreichen. Die Nutzung dieser modernen Alleskönner ist jedoch mit vergleichsweise hohen Kosten (Lizenzen und Rechenzeit) verbunden.

Auch Stimmungen und Einschätzungen zu Technologien können mithilfe von KI aus den Texten herausgelesen und quantifiziert werden. Diese sog. „Sentiments“ können dann als Maß für die Bewertung der Relevanz der Technologien genutzt werden.

1.4. Recherche vs. Vernetzung der Community – oder beides?

Auch ein teilautomatisiertes System benötigt menschliche Expertise, um die ganze Breite an Aufgaben des Technologie-Scoutings abzudecken. Dazu sind neben einer geeigneten Visualisierung auch Funktionalitäten nötig, die die Anreicherung und Korrektur der Informationen sowie die Vernetzung mit Expert/-innen des Sektors ermöglichen.

Informationen in einem TSS sollen leicht erfassbar und analysierbar sein. Hierfür werden in solchen Systemen verschiedene Übersichts- und Detailansichten verwendet, von denen einige auch für den Prototyp in dieser Vorstudie ausgewählt wurden. Ein Knowledge Graph stellt Tech-

nologien und deren Zusammenhänge untereinander sowie zu den anderen Klassen (Produkte, Organisationen, Förderprogramme, Anwendungen) als Netzwerk dar. Ein Trendradar ordnet die Technologien radial nach ausgewählten Metriken (z.B. Reifegrad) an. Diese Übersichts Darstellungen sind für die Exploration wertvoll und können durch Steckbriefsichten mit Detailinformationen ergänzt werden.

Mithilfe von Funktionen der Qualitätssicherung (z. B. ein Hinweis, dass ein Inhalt geprüft wurde) und der Editierung können Nutzende entsprechend ihrer Zugriffsrechte Informationen hinzufügen, korrigieren oder auch löschen. Auf diese Weise kann die ganze User-Community zu einer genaueren und vollständigeren Wissensbasis im TSS beitragen. Ungenauigkeiten der automatisierten Wissens extraktion können korrigiert werden und Nutzende können ihr Wissen, das über die bestehenden Informationen im System hinausgeht, beitragen.

Die Recherche, Einordnung und Kommentierung von Technologien durch die Nutzenden bietet sowohl Mehrwert für die Recherche als auch für die Vernetzung der

Nutzenden untereinander. Die Prominenz von Technologien und Anwendungen im Sektor kann dadurch abgeschätzt werden, wie rege im System nach ihnen gesucht und diese diskutiert werden. Weitere Funktionen wie bspw. die direkte Kontaktaufnahme zu dem Verfasser oder der Verfasserin eines Kommentars oder über ein persönliches Profil können die Vernetzung unterstützen.

In einer Evaluation mit ca. 30 Vertreter/-innen des Schienenverkehrssektors wurde deutlich, dass beide Aspekte wichtig sind – die Überprüfung und hohen Qualitätsstandards der Inhalte genauso wie Beteiligungsformen für die Vernetzung (s. Bild 2). Die beiden Ausrichtungen „Recherchesystem“ und „Community-Plattform“ schließen sich dabei nicht aus. Je nach Schwerpunktsetzung sollten andere Funktionalitäten im Aufbau und der kontinuierlichen Weiterentwicklung priorisiert werden.

Als Verfahrensentwickler möchte ich den Entwicklungsstand von Technologien identifizieren, um die Weiterentwicklung voranzutreiben.

fische Anwendungsfälle aus dem Alltag der Teilnehmenden in Form von „User Stories“ aufgenommen, die zeigen, dass das potenzielle Anwendungsspektrum eines TSS breit ist.

Das Konzept eines teilautomatisierten TSS für den Schienenverkehrssektor wurde in einem Prototyp implementiert, um zu evaluieren, wie gut es sich für die genannten Aufgaben eignet. Die wesentlichen Ziele dabei waren:

1. Test und Evaluation der technischen Eignung von Methoden, Darstellungen und Funktionen
2. Einholen von Feedback aus dem Sektor und Bewertung des Mehrwerts für individuelle Aufgaben

Für den Prototyp wurden sechs verschiedene Publikationsarten auf die urheberrechtlichen Vorgaben für die Nutzung in einem TSS untersucht. Vier Informationsquellen wurden anschließend technisch angebunden.

2. Von der Theorie in die Praxis – Erprobung des Konzepts als Prototyp

In der Evaluation von Vertreter/-innen aus dem Sektor wurde deutlich, dass Rechercheaufgaben derzeit einen wesentlichen Arbeitsaufwand für die Befragten darstellen und Unterstützung hierbei willkommen wäre (s. Bild 2). Über die einfache Recherche hinaus wurden auch zielgruppenspezi-

1 Nutzerin wählt die Trendradaransicht, um sich über Technologien und deren Entwicklungsstand zu informieren

2 Nutzerin setzt die Filter so, dass nur Antriebstechnologien angezeigt werden

3 Nutzerin klickt auf den Marker einer interessanten Technologie, sodass ein Kurzsteckbrief in der Vorschau zu sehen ist

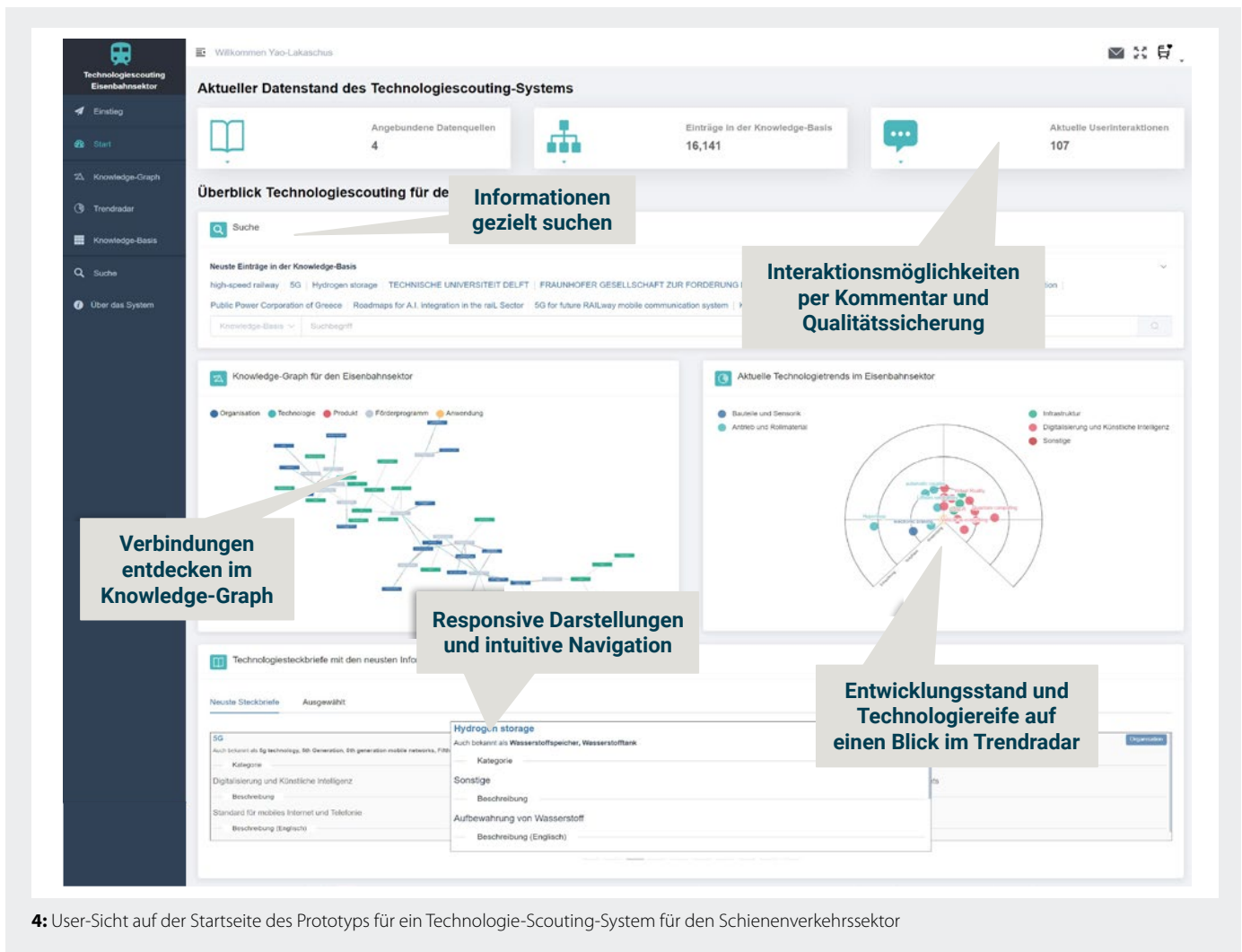
4 Nutzerin möchte weitere Details über die Technologie einsehen und wählt die Steckbriefdarstellung

5 Nutzerin kann dem Steckbrief die gesuchten Informationen zu infrastrukturellen Anforderungen entnehmen und erhält Quellenangaben

6 Nutzerin wechselt zurück zum gefilterten Trendradar, um weitere Technologien im Detail zu recherchieren

3: Illustration der User Story: „Als Infrastrukturplanerin möchte ich den Einsatzhorizont von Antrieben abschätzen, um für den Aufbau der passenden Infrastruktur zu sorgen“

Homepageveröffentlichung unbefristet genehmigt für Deutsches Zentrum für Schienenverkehrsforschung beim Eisenbahn-Bundesamt, d-fine GmbH / Rechte für einzelne Downloads und Auszüge für Besucher der Seiten genehmigt / © DVV Media Group GmbH



4: User-Sicht auf der Startseite des Prototyps für ein Technologie-Scouting-System für den Schienenverkehrssektor

Die Nutzersichten und Navigation wurden anhand der zuvor aufgenommenen User Stories entwickelt (s. Bild 3), wobei sowohl Übersichts- als auch Steckbriefsichten implementiert wurden, um möglichst breites Feedback zu deren Tauglichkeit zu erhalten.

Für die Informationsextraktion wurde ein breites Spektrum an Methoden getestet. Zur Extraktion von Organisationen konnte ein frei verfügbares, fertigtrainiertes Sprachmodell genutzt werden; für die Extraktion von Technologien wurde ein individuelles Modell über maschinelles Lernen trainiert, wobei die Trainingsdaten mithilfe von Technologielisten für den Schienenverkehr automatisch annotiert wurden. Beziehungen zwischen z.B. Technologien und Organisationen wurden regelbasiert über deren gemeinsame Nennung in Texten abgeleitet und für die Kategorisierung von Technologiebegriffen wurde ein großes Sprachmodell (z.B. GPT-3) genutzt. Die erprobten NLP-Verfahren zeigen je nach Aufgabe Unterschiede hinsichtlich Genauigkeit und

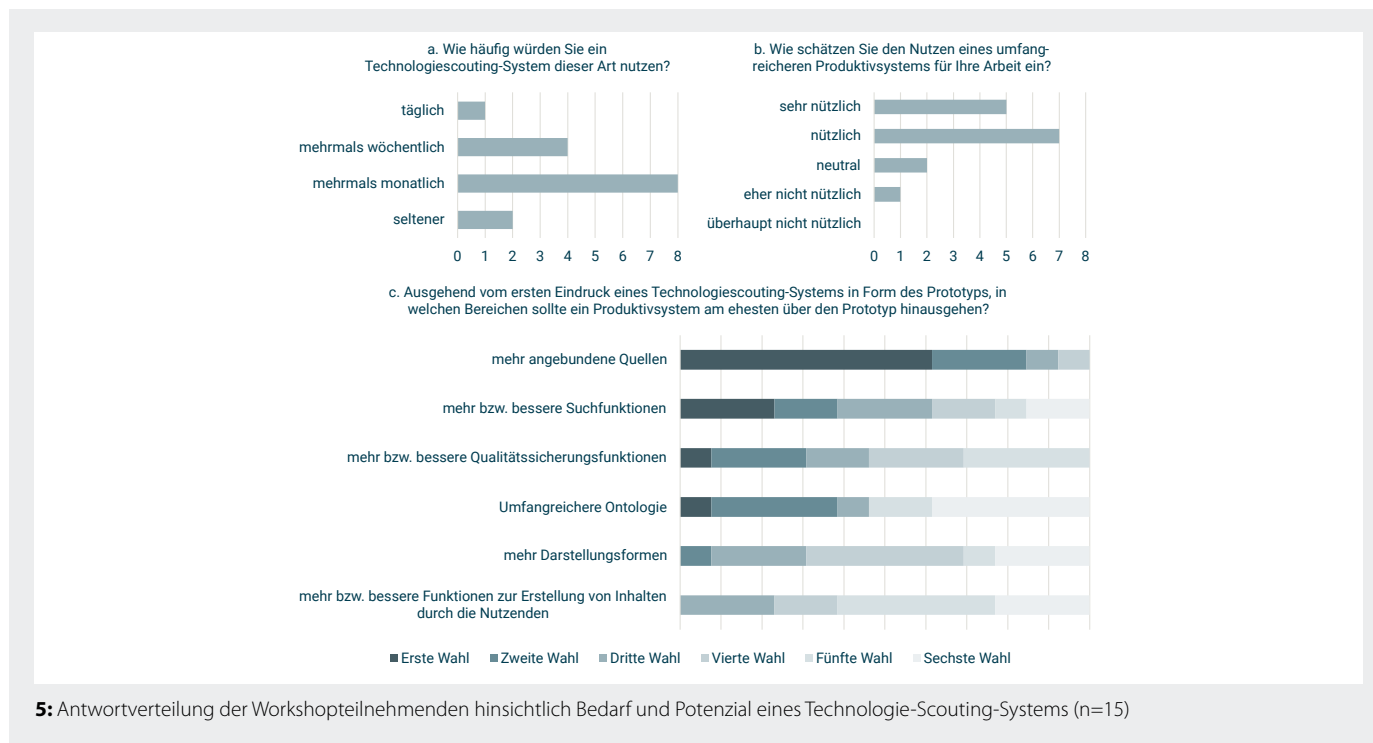
Trefferquote – allerdings sind die aktuellen NLP-Modelle bereits sehr gut für ein TSS geeignet. Insbesondere die Kombination automatisierter Verfahren zur Datenvorverarbeitung mit manueller Korrektur und Qualitätssicherung durch Nutzende (z.B. neue Synonyme derselben Technologie verknüpfen) birgt großes Potenzial.

Technisch beruht der Prototyp auf einer Microservice-Architektur. Die Funktionen der Datenanbindung, -speicherung, Textanalyse und Darstellung sind in isolierten Modulen implementiert, die untereinander über Programmierschnittstellen kommunizieren. Bei der Nutzer-Applikation handelt es sich um eine Web-Anwendung, die im Browser genutzt wird (s. Bild 4).

3. Evaluation mit einer Testgruppe – der Bedarf im Sektor ist da

Der Prototyp wurde mit Vertreter/-innen des Schienenverkehrssektors, anderer Ressortforschungseinrichtungen aus dem

BMDV Expertennetzwerk sowie von Hochschulen anhand von konkreten Anwendungsfällen erprobt. Die Evaluation zeigt, dass ein TSS auf positive Resonanz trifft (s. Bild 5). Eine große Mehrheit gab an, ein solches System mehrmals monatlich oder häufiger nutzen zu wollen. Insbesondere die verschiedenen komplementären Darstellungen (Knowledge Graph, Trendradar, Steckbrief) wurden als hilfreich eingeschätzt. Um seinen Nutzen zu entfalten, müsste ein TSS für den Produktiveinsatz mehr Quellen abdecken, umfangreichere Suchmöglichkeiten bieten und hohe Qualität und Aktualität der Informationen gewährleisten. Qualität, Objektivität und Transparenz bezüglich der verwendeten Methoden und Quellen sind nach Einschätzung der Befragten wichtig für ein hohes Vertrauen in die Informationen. Dies ist nicht nur wesentlich, um die Recherche zu unterstützen, sondern auch, um rege Beteiligung und Motivation der Nutzenden zu erreichen.



4. Vom Test in die Produktion – Betreibermodelle im Vergleich

Im Rahmen der Vorstudie wurde auch untersucht, wie sich Entwicklung und Betrieb eines TSS umsetzen ließen und wer als Betreiber in Frage käme. Dabei wurden verschiedene mögliche Träger identifiziert:

1. ein Konsortium aus Organisationen verschiedener Rechtsformen,
2. ein Unternehmen,
3. eine öffentliche Einrichtung oder
4. eine Forschungsorganisation.

Bei der Bewertung dieser Trägeroptionen wurden die Umsetzbarkeit technischer Aufgaben, organisatorische und strategische Aspekte sowie verschiedene Möglichkeiten der Finanzierung berücksichtigt. Grundsätzlich ist die Umsetzung in allen Varianten möglich, es bestehen für die unterschiedlichen Modelle aber wesentliche Vor- und Nachteile:

- Ein TSS mit dem angedachten Umfang der Recherche und Interaktion für den Schienenverkehrssektor kann sich nicht vollumfänglich auf die urheberrechtliche Schranke für TDM zum Zwecke wissenschaftlicher Forschung stützen.
- In Trägerschaft eines Unternehmens besteht das breiteste Spektrum an Finanzierungsmöglichkeiten. Dabei können

allerdings Varianten wie bspw. Werbung oder Analyse von Nutzerdaten zu einer geringeren Akzeptanz seitens der Nutzenden führen.

- Die Umsetzung als Konsortium erscheint am aussichtsreichsten. Auch wenn mit Abstimmungsaufwänden zwischen Konsortialpartnern zu rechnen ist, bietet sich die Möglichkeit, ein System „vom Schienenverkehrssektor für den Schienenverkehrssektor“ zu schaffen, indem die Interessen verschiedener Stakeholder an oberster Stelle vertreten sind.

5. Fazit

Die Vorstudie hat deutlich gemacht, dass breiter Bedarf nach technischer Unterstützung beim Technologie-Scouting im Schienenverkehrssektor besteht. Die prototypische Umsetzung eines teilautomatisierten TSS hat gezeigt, dass sich intelligente Textanalysemethoden sehr gut für die Wissensextraktion eignen und in Kombination mit menschlicher Expertise eine hohe Informationsqualität erreichen. Die Entwicklung eines Produkktivsystems sollte unter Berücksichtigung der Bereitschaft möglicher Trägerorganisationen weiterverfolgt werden. Als besonders aussichtsreich erscheint dabei die Trägerschaft durch ein Konsortium. Den vollständigen Abschlussbericht zur Vorstudie finden Sie auf der Webseite des DZSF (www.dzsf.bund.de).

Literatur

[1] I. Dincer und C. Zamfirescu, „A review of novel energy options for clean rail applications,“ *Journal of Natural Gas Science and Engineering*, Bd. 28, pp. 461-478, 2016.
 [2] F. O’Sullivan, „Germany Has the World’s First Hydrogen-Powered Passenger Train,“ Available: <http://www.citylab.com/commute/2016/09/germany-hydrogen-passenger-train/501575/>. [Zugriff am 29.09.2021].
 [3] A. Mazzone, M. Schönbacher und X. Larrea, „Future Freight Locomotives in Shift2Rail – Development of Full Electric Last Mile Propulsion System,“ Wien, 2018.
 [4] E. Andersson, O. Fröidh, S. Stichel, T. Bustad und H. Tengstrand, „Green Train: concept and technology overview,“ *International Journal of Rail Transportation*, Bd. 2/1, p. 2.16, 2014.
 [5] A. Ozhigin, P. Pruniv, V. Sverdlin und Y. Vikulina, „Development of an algorithm for energy efficient automated train driving,“ Toulouse, 2016.

Summary

A technology scouting system for research and networking in the railway sector

A semi-automated technology scouting system can support the search for technologies with potential relevance for the rail transport sector. Within the framework of the pilot study presented in this article on the design and implementation of such a system, a technical and functional concept was developed, selected functions were implemented in form of a prototype, an assessment was obtained from experts in the rail sector, and possible operating models were analyzed.