

## Semantische Technologie

Felix Sasaki\*

# FREME – semantische und mehrsprachige Anreicherung selbstgemacht

FREME erlaubt die semantische Anreicherung von digitalen Inhalten, für und von Wissensarbeitern adaptierbar hinsichtlich inhaltlicher Domänen, Datenformate und Anwendungsszenarien

<https://doi.org/10.1515/iwp-2017-0055>

## Motivation für FREME

Das Thema „Künstliche Intelligenz“ ist in aller Munde. Nahezu tagtäglich hören wir von neuen Superlativen, die mit den Methoden der KI in unterschiedlichen Lebens- und auch Wirtschaftsbereichen erzielt werden. Größtenteils sind diese Methoden nur den Experten zugänglich. Man muss sich mit den Verfahren und Werkzeugen der KI auskennen, um Mehrwerte zu schaffen. Das Projekt FREME hat sich zum Ziel gesetzt, dies für einen ausgewählten Bereich der KI zu ändern: Die semantische und mehrsprachige Anreicherung digitaler Inhalte.

Unter „digitale Inhalte“ verstehen wir hier unstrukturierte Daten z. B. in textuellen Dokumenten, oder teilstrukturierte Daten, z. B. Metadaten. „Semantische und mehrsprachige Anreicherung“ meint die Verknüpfung der Inhalte mit zusätzlichen Informationen, teilweise über Sprachgrenzen hinweg. Dabei kommen Tools zum Einsatz, mit denen viele von uns in unterschiedlichen Kontexten schon länger arbeiten, z. B. zur automatischen Übersetzung. FREME möchte es Wissensarbeitern erleichtern, mit diesen Tools umzugehen und sie zu adaptieren. Ein zentraler Mehrwert von FREME ist, dass die Wissensarbeiter diese Adaption vornehmen können, ohne die zu Grunde liegenden Verfahren der KI im Detail zu kennen.

## FREME: Geschäftsfelder und Anwendungsszenarien

FREME stellt eine Open Source technische Infrastruktur, d. h. ein Framework, bereit, das den Zugriff auf Anreicherungstechnologien und die Integration in bestehende Prozesse erleichtert. Das FREME Framework ist unter der Lizenz Apache 2.0 unter <https://freme-project.github.io/> verfügbar. Es wurde in einem EU Projekt zwischen Februar 2015 und Januar 2017 entwickelt und vor allem durch vier Geschäftsfelder und Anwendungsszenarien vorangetrieben:

1. Erstellung und Veröffentlichung mehrsprachiger und semantisch angereicherter eBooks;
2. Integration von Anreicherungen in Übersetzungsworkflows;
3. Erleichterung zu digitalen Inhalten im Bereich Ernährungsforschung durch Anreicherung von Metadaten; sowie
4. FREME unterstützte, personalisierte Empfehlung von Inhalten.

Mit Absicht wurde eine große Bandbreite an Szenarien gewählt, um Feedback aus möglichst vielen Geschäftsfeldern sammeln zu können. Eine wichtige Erfahrung am Ende des Projektes war, dass KI-Technologien nicht per se (wirtschaftliche) Mehrwerte mit sich bringen. Man muss im Detail in Branchen schauen, wo sich der Einsatz unter welchen technischen und teilweise nicht technischen Voraussetzungen wirklich lohnt. Das wollen wir nun am Beispiel mehrsprachiger eBooks tun.

## Mehrwehrt für die Verlagsbranche

Unsere Diskussion mit der Verlagsbranche hat ein großes Interesse im eLearning Bereich gezeigt. Die Anreicherung von eBooks für Sprachlerner ist ein relevantes Szenario für

---

\*Kontaktperson: Dr. Felix Sasaki, LambdaWerk GmbH,  
Oranienburger Straße 87, 10178 Berlin,  
E-Mail: felix.sasaki@lambdawerk.com

## FREME API Documentation

This section only covers the most important endpoints of FREME: the enrichment endpoints.

The endpoints can be used to access FREME e-Services via common HTTP requests.

A full documentation of all e-Service endpoints, including all parameters, is provided [here](#). For usage examples, see the [tutorial section](#).

POST	/e-entity/dbpedia-spotlight/documents	Entity recognition and linking using DBpedia Spotlight engine.
POST	/e-entity/freme-ner/documents	Entity recognition and linking using Freme-NER engine.
POST	/e-link/explore	Retrieve Description for a resource
POST	/e-link/documents/	Fetch data about named entities from various ontologies
POST	/e-publishing/html	Create eBooks in the EPUB3 format
POST	/e-terminology/tilde	Annotate text with terminology information using Tilde Terminology service.
POST	/e-translation/tilde	Annotate text with terminology information using Tilde Terminology service.

Abbildung 1: Screenshot Überblick zu den FREME Programmierschnittstellen (APIs).

entsprechende Verlage. Wichtig ist aber die Adaptierbarkeit der Anreicherung. Viele semantische Tools erzeugen Links auf Wikipedia Einträge. Das ist für Sprachlernen nicht hilfreich. Der Verlag muss in der Lage sein, eigene Ressourcen einzusetzen und z. B. Links auf Übersetzungsvorschläge in mehrsprachigen Lexika zu setzen.

Neben der Adaptierbarkeit hat dieses Anwendungsszenario deutlich gemacht, dass die Unterstützung von Standards von großer Bedeutung ist. Ohne diese Unterstützung besteht die Gefahr, Anreicherung nur auf einem eBook Lesegerät nutzen zu können. FREME unterstützt deshalb Formate wie HTML5 und ePub, um die Nutzung in Lesegeräten zu erleichtern. Des Weiteren kommen im FREME Framework selbst Standards zum Einsatz, um Anreicherungsinformationen zu speichern, zu kombinieren und in verschiedenen Workflows zu integrieren. Genannt werden sollen hier der Linked Data Technologiestack, die Linked Data Anwendungen NIF bzw. OntoLex / Lemon, sowie das Internationalization Tag Set (ITS) 2.0.

Die Fülle dieser Standards ist erschlagend. Das Anwendungsszenario digitales Publizieren und auch die anderen Geschäftsfelder haben gezeigt, dass die Details der Standards für den Nutzer irrelevant sind. Deshalb verwendet FREME die standardisierten Technologien intern, um Anreicherungsworkflows zu konfigurieren. Für die Anwender, z. B. Wissensarbeiter im Verlag für Lerninhalte, wird FREME durch graphische Schnittstellen bereitgestellt. Hier kann sie oder er die Anreicherung steuern, z. B. ein Lexikon als Quelle für die Anreicherung auswählen, und die Ergebnisse nachbearbeiten um die notwendige Qualität zu erhalten.

Die Integration von Anreicherungstools ist oft mit dem Aufwand der Formatanpassung verbunden. FREME hat für die Verlagsbranche im Bereich „technische Dokumentation“ Unterstützung anzubieten, z. B. für das gebräuchliche DocBook Format. Die XML-Unterstützung von FREME erleichtert die Nutzung für viele Verlage, da XML hier eine erhebliche Verbreitung besitzt.

Die Geschäftsfelder haben im Projekt auch gezeigt, dass die Notwendigkeit zur Nachbearbeitung vom Szenario abhängt. Z. B. für den Lernbereich ist es unabdingbar, keine Fehler in der Anreicherung zu haben. Deshalb muss ein Wissensarbeiter die automatischen Ergebnisse korrigieren können, ohne dabei in die KI-Technologien tief einzusteigen.

## Mehrwerte für Wissensarbeiter

Die beschriebenen Szenarien aus der Verlagswelt sind nur ein Beispiel für die Nutzung von FREME. Ein anderes Beispiel ist die Nutzung in Wissensseinrichtungen wie Bibliotheken. Hier ist es eine häufige Aufgabe, Metadaten zu Disambiguieren und eine semantische Suche zu ermöglichen. FREME erlaubt es entsprechende Datenquellen wie ORCID zu konfigurieren und XML basierte Metadatenformate zu verarbeiten.

Wir haben in diesem Artikel schon öfter den Begriff „Wissensarbeiter“ verwendet. Was Wissensarbeiter sind und wie sie FREME am besten nutzen, lässt sich nicht generalisieren. Über Anwendungsbereiche hinweg hat das Projekt aber gezeigt, dass die KI-Technologie der semanti-

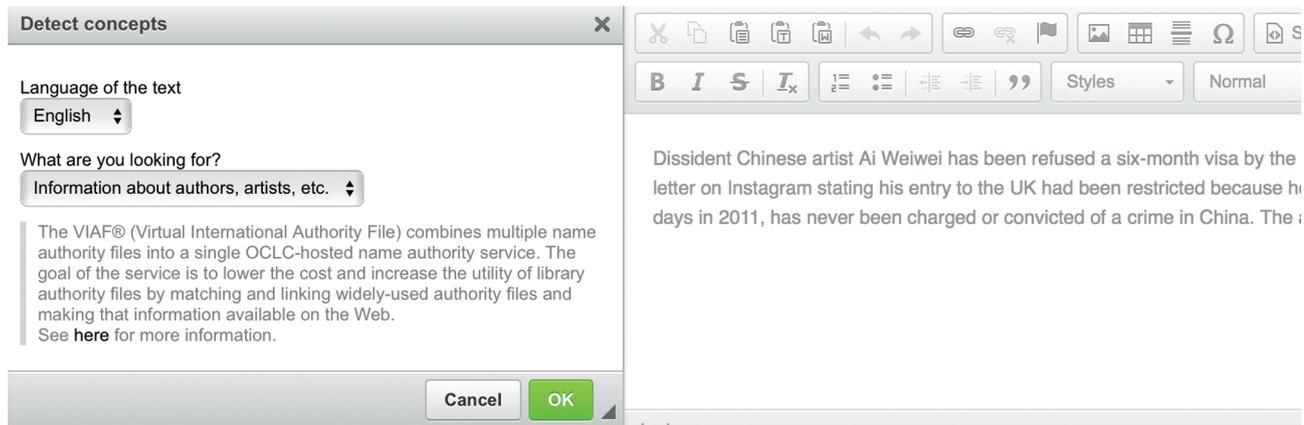


Abbildung 2: Screenshot Nutzung der VIAF Datenquelle in einem Browser basierten Editor.

schen und mehrsprachigen Anreicherung nur mit Wissensarbeitern echte Mehrwerte schafft. Das beschriebene Beispiel aus dem Bereich hat nur eine Rolle beschrieben: die Nachbearbeitung von Anreicherungsergebnissen. Wissensarbeiter haben aber viele andere Aufgaben, die vor allem mit Datenquellen für die Anreicherung zu tun haben, vgl. etwa die erwähnte Metadatenverarbeitung. Dabei geht es unter anderem um die folgenden Aspekte:

- Qualität: sind die Daten derart mit Fehlern behaftet, dass sie dem jeweiligen Anwendungsszenario nicht genügen?
- Aktualität: steht ein Datensatz zur Verfügung der dem Aktualitätsanspruch genügt?
- Verlässlichkeit: sind die Fakten, welche in den Daten wiedergegeben werden, verlässlich?
- Verfügbarkeit: stehen die Daten in einer Form zur Verfügung, welche die kommerzielle Nachnutzung ermöglicht?
- Stabilität: sind die Daten nachhaltig verfügbar?

Die Erfahrungen in FREME haben gezeigt, dass es oft derartige, teilweise nicht technische Aspekte sind, welche die reale Nutzung von KI-Technologien aus dem Bereich semantische und mehrsprachige Anreicherung digitaler Inhalte behindern.

## Nächste Schritte

Das FREME Framework wird von den Projektpartnern inzwischen kommerziell genutzt. Zudem steht eine Live-Installation zur Verfügung, die das ADAPT Centre in Irland

bereitstellt. Bereits vor Projektende hat sich gezeigt, dass ein technisches Rahmenwerk nicht allein ausschlagend ist. Gebraucht werden Nutzer, die sich für die beschriebenen und anderen Szenarien interessieren und die Input für die Fortführung des Frameworks geben. Der am Anfang dieses Artikels gegebene Link zeigt auch den Zugang zu dieser Community, die über die Projektgrenzen hinausgeht. Es ist meine Hoffnung, dass sich auch Wissensarbeiter aus den unterschiedlichsten Bereichen für FREME interessieren und ihre Kompetenz einbringen. Nur so können echte Mehrwerte bei der Anreicherung digitaler Inhalte entstehen.

**Deskriptoren:** Computerlinguistik, Maschinelles Indexierungsverfahren, Systemumgebung, Projekt



**Dr. Felix Sasaki**  
 LambdaWerk GmbH  
 Oranienburger Straße 87  
 10178 Berlin  
 felix.sasaki@lambdawerk.com

Dr. Felix Sasaki hat im W3C zwischen 2005 und 2009 im Bereich Internationalisierung gearbeitet. Ab 2012 hat er als W3C Fellow für das DFKI (Deutsches Forschungszentrum für künstliche Intelligenz) die MultilingualWeb-LT Working Group mitgeleitet. Er ist Co-Editor der Internationalization Tag Set (ITS) 2.0 Spezifikation. Sein Hauptinteresse gilt der Nutzung von Webtechnologien für die Repräsentation und Verarbeitung mehrsprachiger Daten.