

Schwerpunkt Open Science

Christin Seifert und Michael Granitzer, Passau

EEXCESS: Personalisierter Zugriff auf Long-Tail-Inhalte – neue Methoden zur Verbreitung wissenschaftlich-kulturellen Wissens

Traditionell bieten wissenschaftliche und kulturelle digitale Bibliotheken Zugriff auf ihre Inhalte über eigene Portale und dedizierte Suchmechanismen an. Ein wichtiger Aspekt in der Verbreitung des darin enthaltenen Wissens ist die Bewerbung dieser Dienste und die Gewinnung neuer Nutzer und Nutzerinnen auf den Portalen. Dabei kommen vor allem Methoden aus dem Bereich der Suchmaschinenoptimierung und des Social Media-Marketings zum Einsatz. Betrachtet man jedoch die Gesamtheit der digitalen Bibliotheken, so zeigt sich, dass diese Mittel nur begrenzt zum Erfolg führen können. Wissenschaftliche und kulturelle digitale Bibliotheken, Museen und Archive stellen im WWW den so genannten „Long-Tail“ der Inhalte dar, d. h. die große Menge hoch-spezialisierter Information für eine auf die einzelnen Inhalte bezogene sehr kleine Nutzergemeinschaft. Diese Long-Tail-Inhalte konkurrieren mit Mainstream-Inhalten um die Gunst von Suchmaschinen und Social Media-Kanälen. Die Treffsicherheit der Methoden liegt daher meist unter den Erwartungen. Das EU FP 7-Projekt EEXCESS untersucht neue Möglichkeiten zur Verbreitung wissenschaftlich-kultureller Inhalte digitaler Bibliotheken. Im Zentrum stehen dabei intelligente, die Privatsphäre erhaltende, personalisierte Empfehlungstechniken für wissenschaftliche und kulturelle Long-Tail-Inhalte. Diese Inhalte sollen dabei in häufig genutzte Web-Kanäle automatisiert eingespeist und somit einfacher nutzbar und für eine breitere Anwenderbasis sichtbar gemacht werden. EEXCESS bringt die Inhalte zu den Nutzern und Nutzerinnen, anstatt Nutzer und Nutzerinnen zu den Inhalten zu führen – so das Ziel.

Deskriptoren: Elektronische Bibliothek, Empfehlung, Visualisierung, Kontext, Projekt, EEXCESS

EEXCESS: Personalised access to Long Tail content – New strategies for disseminating scientific and cultural content

In the last decade, Europe has conducted a tremendous effort to make cultural, educational and scientific resources publicly available. Although such massive amounts of culturally and scientifically rich content are available, the potential of its use for educational and scientific purposes remains largely untapped. One reason can be seen in current web content dissemination mechanism, which are dominated by a small number of large central hubs like major search engines (e.g. Google), social networks (e.g. Facebook) or online encyclopaedias (e.g. Wikipedia). In order to maintain their valuable services, those large hubs have to focus on commercially viable mainstream content. While cultural and scientific resources provide valuable and educational content, they cannot be considered as ‘mainstream’. Quite contrary, most of this can be considered as high-quality niche content for a rather small community and forms part of the so-called Long Tail. The Long Tail theory, first introduced by Chris Anderson, argues that in internet-based markets, niche content adds up to a huge body of knowledge, but is hidden from most users. In the Long Tail, content is maintained and curated by a large number of small to medium-sized institutions such as memory organisations (e.g. archives and museums), national and digital libraries and open educational repositories. However, the few large web hubs hardly support the dissemination of this Long Tail content leaving a gap for bringing cultural and scientific wealth into educational and scientific processes. In order to reshape content dissemination mechanisms for highly specialised Long Tail content, EEXCESS introduces the idea on augmenting existing web channels with high quality content through personalised, contextualised and privacy preserving recommendations. In order to communicate the knowledge contained in the content, EEXCESS researches visualisation and interaction techniques for presenting recommendation results.

The main underlying concept is to bring the content to the user, i.e. injecting content into channels used by users, instead of bringing the user to the content, i.e. creating additional portals that compete for user attention in the Long Tail.

Descriptors: Electronic Library, Recommendation, Visualization, Context, Project, EEXCESS

EEXCESS: Accès personnalisé au contenu « long tail » – de nouvelles méthodes pour diffuser les connaissances scientifiques et culturelles

Traditionnellement, les bibliothèques numériques scientifiques et culturelles offrent un accès à leur contenu par le biais de leurs propres portails et moteurs de recherche dédiés. La publicité autour de ces services revêt un aspect important dans la diffusion de ces connaissances et ce afin d'attirer de nouveaux utilisateurs et utilisatrices sur les portails. Pour y parvenir, l'optimisation des moteurs de recherche et plus récemment, le marketing des médias sociaux sont essentiellement utilisés. Cependant, si l'on considère l'ensemble des bibliothèques numériques, force est de constater que ces moyens ne peuvent mener qu'à un succès limité. Les bibliothèques numériques scientifiques et culturelles forment dans l'internet le soi-disant contenu « long tail », c.-à-d. une grande quantité d'informations hautement spécialisées pour une communauté d'utilisateurs relativement petite. Ce contenu « long tail » doit concurrencer les contenus grand public pour attirer l'attention des moteurs de recherche et des médias sociaux. La précision des méthodes est généralement en-deçà des attentes. Le projet européen FP 7 EEXCESS examine de nouveaux moyens de diffusion du contenu scientifique et culturel des bibliothèques numériques. Ce projet met l'accent sur des techniques de recommandation intelligentes et personnalisés qui préservent la vie privée. Les contenus devraient alimenter automatiquement des canaux web qui sont utilisés de façon régulière, ce qui permettrait de les rendre plus faciles à utiliser et visibles à un public plus large. En d'autres termes, le projet EEXCESS vise à apporter le contenu aux utilisateurs, au lieu d'essayer d'attirer l'utilisateur vers le contenu.

Describeurs: Bibliothèque numérique, Recommandation, Visualisation, Contexte, Projet, EEXCESS

DOI 10.1515/iwp-2015-0022

1 Einleitung

In der letzten Dekade wurden enorme Anstrengungen unternommen, um kulturelle und wissenschaftliche Inhalte digital zur Verfügung zu stellen. Trotz der grundlegend hohen Verfügbarkeit der Inhalte bleibt deren Potenzial zum Großteil unerschlossen. Dies ist einerseits auf die Art und Eigenschaften der Inhalte zurückzuführen und andererseits auf die im Web dominanten Such- und Findemechanismen.

Der vorliegende Artikel stellt das im 7. Rahmenprogramm der EU geförderte Projekt EEXCESS¹ vor, das mit zehn europäischen Forschungs- und Anwendungspartnern aus fünf Ländern neue Methoden zur Verbreitung, Verknüpfung und Präsentation kultureller und wissenschaftlicher Inhalte entwickelt und als Open-Source-Lösung zur Verfügung stellt. EEXCESS wurde im Februar 2013 gestartet und hat eine Projektlaufzeit von 3,5 Jahren. Ausgangspunkt stellt dabei die Betrachtung der im Web dominanten Such- und Findeprozesse dar. Aktuelle webbasierte Such- und Findeprozesse bedienen sich einer kleinen Anzahl an zentraler Diensten, wie z. B. Suchmaschinen (z. B. Google), Soziale Netzwerke (z. B. Facebook, Twitter) oder kollaborativ erstellte Inhalte (z. B. Wikipedia). Aus verschiedenen Gründen fokussieren diese wenigen zentralen Dienste auf sogenannte Mainstream-Inhalte, also jene geringe Menge an Inhalten, welche viele Nutzerinnen und Nutzer interessieren. Wissenschaftlich-kulturelle Inhalte fallen nicht in diese Kategorie. Im Gegenteil, sie stellen sogar eine große Menge hoch-spezialisierter, hoch-qualitativer Inhalte dar, die im einzelnen nur für eine überschaubare Anzahl von Nutzerinnen und Nutzer von Interesse ist. Sie bilden somit den sogenannten Long-Tail der Inhalte.

Hier setzt EEXCESS an und stellt die Frage, wie die von vielen verschiedenen Institutionen bereitgestellten, wissenschaftlich-kulturellen Long-Tail-Inhalte den interessierten Nutzerinnen und Nutzern zu Verfügung gestellt werden können. Ziel ist es dabei, nicht die Nutzen zu den Inhalten zu bringen, sondern die Inhalte zu den Nutzenden. Durch neu entwickelte Technologien sollen dabei Inhalte in bestehende Web-Kanäle möglichst zielgenau eingebracht werden. Dazu ist es notwendig, über personalisierte Empfehlungsmechanismen bestehende Web-Inhalte mit kulturellen-wissenschaftlichen Inhalten anzureichern und unaufdringlich zu präsentieren.

¹ Enhancing Europe's eXchange in Cultural, Educational and Scientific Resources

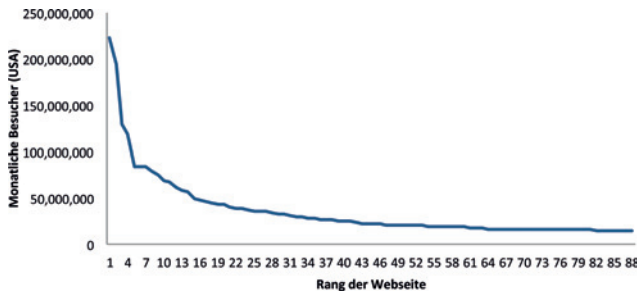


Abb. 1: Beispiel einer Long-Tail-Verteilung anhand der monatlichen Besucher pro Webseiten absteigend sortiert (basierend auf Statistiken von Quantcast²).

Im Folgenden diskutieren wir die Details dieses Ansatzes beginnend bei einer Analyse der Eigenschaften von Long-Tail-Inhalten (siehe Abschnitt 2). Darauf aufbauend geben wir einen Überblick über die Forschungsziele von EEXCESS in Abschnitt 3 und beleuchten anschließend jedes Forschungsziel im Detail (Abschnitte 3.1 bis 3.4), um in Abschnitt 4 Resümee zu ziehen.

2 Long-Tail-Inhalte: Mehrwert und Herausforderungen

Der Begriff des Long-Tails wurde von Chris Anderson, Editor-in-Chief des Wired Magazins, im Kontext des Web 2.0 geprägt (Anderson 2004). Er charakterisiert sich über die Verteilung der Popularität von Produkten und die Anzahl der Produkte selbst. Betrachtet man diese Verteilung, so sieht man, dass wenige Produkte eine sehr hohe Popularität aufweisen, während viele Produkte unpopulär sind. Diese unpopulären Produkte formen den sogenannten Long-Tail, d. h. das lange Ende der Verteilung. Diese Verteilung trifft nicht nur auf Produkte und deren Popularität zu, sondern auch auf (Web) Inhalte und deren Nutzung. Abbildung 1 zeigt eine solche Long-Tail-Verteilung über die Anzahl der monatlichen Besucher und Besucherinnen von Webseiten.

Insbesondere verteilen sich wissenschaftliche und kulturelle Inhalte meist über eine Vielzahl von unterschiedlich spezialisierten Institutionen, wie z. B. Bibliotheken, Museen oder Archive. Die Schwierigkeit der Nutzung der Inhalte im Long-Tail kann dabei auf ihre Verteilung, Diversität, Größe und Spezialisierung zurückgeführt werden. Diese Eigenschaften definieren aber auch den Mehrwert des Long-Tails: die Entdeckung von neuem Wissen oder die Validierung von bestehenden In-

halten im Web. Während z. B. populärwissenschaftliche Artikel ein Thema nur oberflächlich abhandeln, ermöglicht die entsprechende wissenschaftliche Literatur, das Thema im Detail zu erörtern und Zusammenhänge zu validieren. Ähnlich ermöglichen kulturelle Inhalte die meist multimediale Betrachtung wertvoller Artefakte und historischer Ereignisse und führen, so wie wissenschaftliche Inhalte, oft zur Entdeckung neuer Zusammenhänge. Aufgrund des breiten Spektrums an Inhalten, gilt dies für nahezu alle Themengebiete.

Derzeitige Such- und Findprozesse sind aus zwei Gründen suboptimal zur Nutzbarmachung dieser wertvollen Long-Tail-Inhalte:

1. Long-Tail-Inhalte konkurrieren mit den populärsten Inhalten um erste Plätze bei Suchmaschinen-Rankings oder um die Aufmerksamkeit der Nutzer und Nutzerinnen in Social Media. Hinter der Optimierung dieser Rankings stehen meist kommerzielle Interessen (vgl. Suchmaschinenoptimierung und Social Media-Marketing), mit denen Bibliotheken, Museen und Archive nur schwer konkurrieren können.
2. Long-Tail-Inhalte benötigen einen passenden Nutzungskontext, um ihren vollen Wert zu entfalten. Ähnlich einem sehr guten wissenschaftlichen Überblicksartikel, der verschiedene Arbeiten miteinander in Verbindung setzt, benötigen Long-Tail-Inhalte eine ähnliche Klammer, reflektiert entweder über das Wissen der Nutzerinnen oder des Nutzers, oder aber in Verbindung mit weiterführender Information.

Daraus ergibt sich die Frage, wie diese Kontextualisierung von Long-Tail-Inhalten technologisch unterstützt werden kann. Eine Frage, die EEXCESS zu beantworten versucht.

3 EEXCESS – Enhancing Europe’s eXchange in Cultural, Educational and Scientific Resources

Das Ziel von EEXCESS³ ist die Kontextualisierung von Long-Tail-Inhalten durch die Bereitstellung dieser Inhalte in von Nutzern und Nutzerinnen regelmäßig genutzten Web-Kanälen. Die Web-Kanäle werden dabei mit hochqualitativen wissenschaftlichen und kulturellen Inhalten angereichert (augmentiert), welche idealerweise sowohl zum aktuellen Inhalt des Web-Kanals als auch zur Erwar-

² <http://www.quantcast.com/top-sites/US> [01.03.2015]

³ <http://eexcess.eu>

tung und zum Wissenstand der Benutzer und Benutzerinnen passen.

Um diese anspruchsvolle Aufgabe zu lösen, müssen folgende Fragen beantwortet werden:

1. Welche Kanäle sollen angereichert werden, um möglichst treffsicher das Zielpublikum zu erreichen?
2. Wie kann der Kontext für Long-Tail-Inhalte, sowohl durch Analyse des Nutzerverhaltens als auch durch Analyse des Inhalts des Web-Kanals ermittelt werden?
3. Wie können Inhalte unter Nutzung des ermittelten Kontexts im Long-Tail gefunden und bereitgestellt werden?
4. Wie können Zusammenhänge von Long-Tail-Inhalten dem Publikum präsentiert werden?

In den nachfolgenden Abschnitten beschreiben wir die aktuellen Forschungs- und Entwicklungsarbeiten zur Beantwortung dieser Fragestellungen im Projekt EEXCESS. Die technologische Realisierung steht – im Sinne der Offenheit – als Open-Source-Lösung zu Verfügung⁴.

3.1 Identifikation geeigneter Web-Kanäle

Für die Verteilung von Long-Tail-Inhalten an potenziell interessierte Nutzergruppen ist es wichtig zu wissen, in welchen Web-Kanälen sich diese aufhalten. Im WWW gibt es einige wenige Webseiten, die sehr stark verlinkt sind und als Information-Hubs für andere weniger stark verlinkte Teile des Webs fungieren (Barabasi et al. 2000). Diese Information-Hubs repräsentieren gleichzeitig die Webseiten, die am häufigsten aufgerufen werden, als Beispiele wären Google, Wikipedia und die Blogging Plattform Blogger zu nennen.

Eine Möglichkeit um ein breites Publikum zu erreichen, ist es daher, die Inhalte in solche Information Hubs zu integrieren.

Einige dieser Hubs, wie z. B. Blogger oder Wordpress, bieten Plugin-Funktionalität an, um die Plattformen funktional zu erweitern. Andere wiederum etablieren ein vollständiges digitales Ökosystem. Als Beispiel wäre der Google App Market für Googles webbasierte Office-Lösungen zu nennen. Für Webseiten, bei denen eine direkte Einbindung von Inhalten nicht möglich ist, können der Nutzer oder die Nutzerin auf Ebene des Browsers unter Verwendung sogenannter Extensions unterstützt werden (Schlötterer et al. 2014). Technologisch

basieren dabei alle diese Plattformen auf Webtechnologien und -standards. Durch größtmögliche Modularisierung der Architektur, Verwendung von Webtechnologien und -standards (HTML, CSS, JavaScript) kann ein breites Spektrum an Zielplattformen mit minimalem Portierungsaufwand abgedeckt werden. EEXCESS basiert auf diesen Technologien und unterstützt derzeit Wordpress, Google Docs, den Web Browser Chrome sowie spezialisierte Lern-Management Systeme. In der Zukunft sind Erweiterungen für Firefox und Moodle – ein weit verbreitetes offenes Lern-Management-System – geplant.

Eine weitere Charakterisierung von Kanälen ergibt sich über die Nutzungsart von Inhalten und den beteiligten Inhaltsverteilungsprozessen. Hier können zwei prinzipielle Szenarien unterschieden werden: (a) der Konsum von Inhalten und (b) die Erstellung von Inhalten (Granitzer et al. 2013).

Charakteristisch für die Nutzung von Long-Tail-Inhalten beim Konsum ist zum einen der Wunsch nach Belegbarkeit von Inhalten im aktuellen Kontext und zum anderen die Möglichkeit, neue Zusammenhänge zu entdecken, die das Gesamtwissen ergänzen. Eine konkrete Instanziierung dieses Konsumszenarios wäre eine automatische Verlinkung von Webseiteninhalten mit Long-Tail-Inhalten.

Spezifisch für die Erstellung von Inhalten ist hingegen ihre Erweiterung und Anreicherung mit Long-Tail-Inhalten. Ein Beispiel dafür sind Empfehlungen von relevanten Long-Tail-Inhalten beim Verfassen eines Blogbeitrags und die Einbindung einer passenden Graphik oder Referenz in den Blogbeitrag. Letztgenanntes Szenario impliziert automatisch ein Konsumszenario für die Leser und Leserinnen des Blogbeitrags, womit hier Multiplikatoren-Effekte erreicht werden können. Abbildung 2 zeigt am Beispiel der kollaborativen Textverarbeitung „Google Docs“, wie sich Long-Tail-Inhalte in Inhaltserstellungsszenarien integrieren lassen.

3.2 Bestimmung des Kontexts

Wie beschrieben, liegt ein Mehrwert von Long-Tail-Inhalten in der kontextualisierten Anreicherung von Web-Kanälen, sowohl in Konsum- als auch in Erstellungsprozessen. Die zentrale Frage hierbei ist, was diesen Kontext ausmacht bzw. woraus sich dieser Kontext zusammensetzt.

In unserem Fall betrachten wir den Kontext als konkretes, nicht notwendigerweise expliziertes Informationsbedürfnis eines Nutzers oder einer Nutzerin. Dieses umfasst (a) die Detektion „ob“ ein Informationsbedürfnis vorliegt und (b) um „welches“ Informationsbedürfnis es

⁴ <http://github.com/EEXCESS/eexcess>

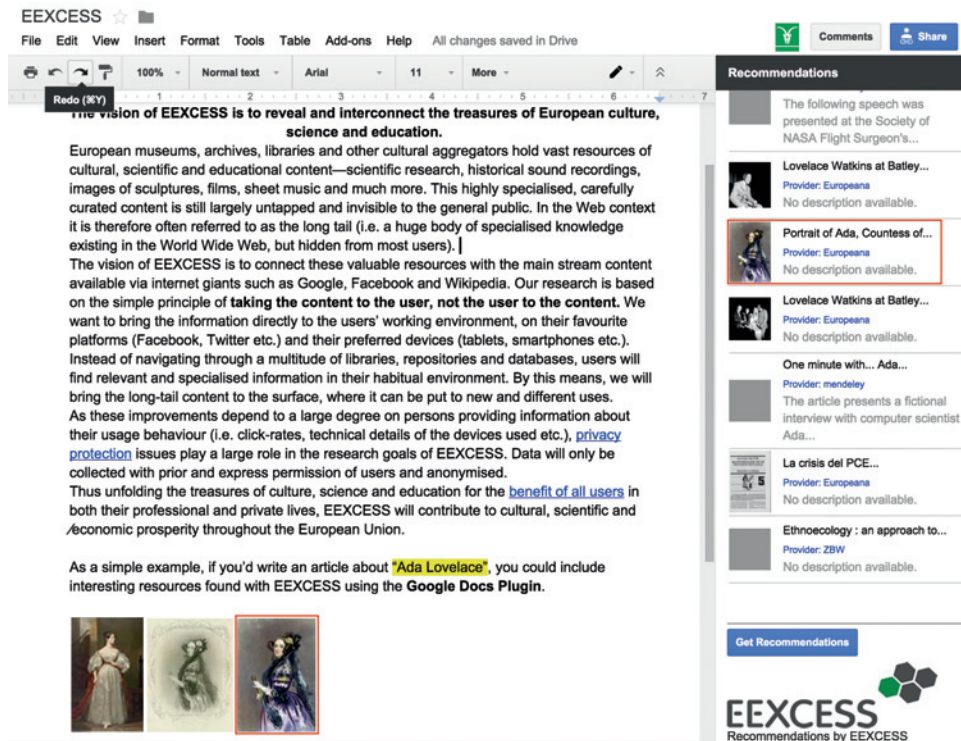


Abb. 2: EEXCESS Plugin für Google Docs am Beispiel. Rechts erhält der Benutzer Vorschläge zum aktuell geschriebenen Absatz mit der Möglichkeit Inhalte und/oder Referenzen zu übernehmen.

sich handelt. Wesentliche Einflussfaktoren dafür sind einerseits die aktuell betrachteten Inhalte (z. B. Webseite, Paragraph) und andererseits das Nutzerprofil, d. h. eine maschinenlesbare Repräsentation einer Person zum Zweck der Identifikation der Person und der Personalisierung von Inhalten (Carberry et al. 2013). Nutzerprofile können durch explizite Angaben der Nutzenden oder aber durch längere Beobachtung der betrachteten Inhalte und Interaktionen aufgebaut werden. Ziel einer solchen Personalisierung ist die Präsentation von „richtigen“ Inhalten zur „richtigen“ Zeit in der „richtigen“ Art und Weise (Fischer 2001).

In EEXCESS bedeutet dies, dass aufgrund eines erstellten Nutzerprofils die richtigen Long-Tail-Inhalte gefunden und in die entsprechenden Kanäle in ansprechender Art und Weise eingespeist werden.

Konzeptuell lassen sich vier verschiedene Granularitätsstufen zur Ableitung des aktuellen Kontextes definieren: Phrasen, Paragraphen, Webseiten und Sessions.

Phrasen: Der Informationsbedarf ist über eine Phrase bestimmt, diese kann z. B. durch eine Verlinkung oder eine aktive Selektion des Nutzers oder der Nutzerin definiert sein. Dieser Informationsbedarf ist sehr spezifisch, und auch die relevanten Inhalte sind sehr spezifischer Natur. Unsere Experimente haben dabei gezeigt, dass

sich ein Informationsbedürfnis auf Phrasen-Ebene mit hoher Genauigkeit (80%) automatisiert voraussagen lässt (Seifert et al. 2015).

Paragraph: Ein kompletter Paragraph bestimmt hier den Informationsbedarf. Paragraphen behandeln im Allgemeinen ein abgeschlossenes Thema, das über den Text automatisch erschlossen werden kann.

Webseite: Der Informationsbedarf ist auf der Ebene der Webseite bestimmt, die als Sequenz von Paragraphen gesehen werden kann. Der Informationsbedarf auf Webseitenebene ist thematisch breiter als auf den vorgenannten Granularitätsstufen und durch allgemeinere Inhalte abdeckbar.

Session: Auch eine komplette Suchsession kann den Informationsbedarf definieren, z. B. wenn Nutzer und Nutzerinnen auf verschiedenen Seiten zu einem bestimmten Themenkomplex recherchieren. Außerdem lassen sich durch Beobachtung über einen längeren Zeitraum Interessensprofile ableiten, die wiederum für die Spezifizierung des Informationsbedarfes auf anderen Granularitätsstufen verwendet werden können.

EEXCESS nutzt diese Granularitätsstufen um Benutzerprofile aufzubauen und zielgerichtet kulturelle und wissenschaftliche Long-Tail-Inhalte in hochfrequente Web-Kanäle und Inhaltsverteilungsprozesse zu integrieren.

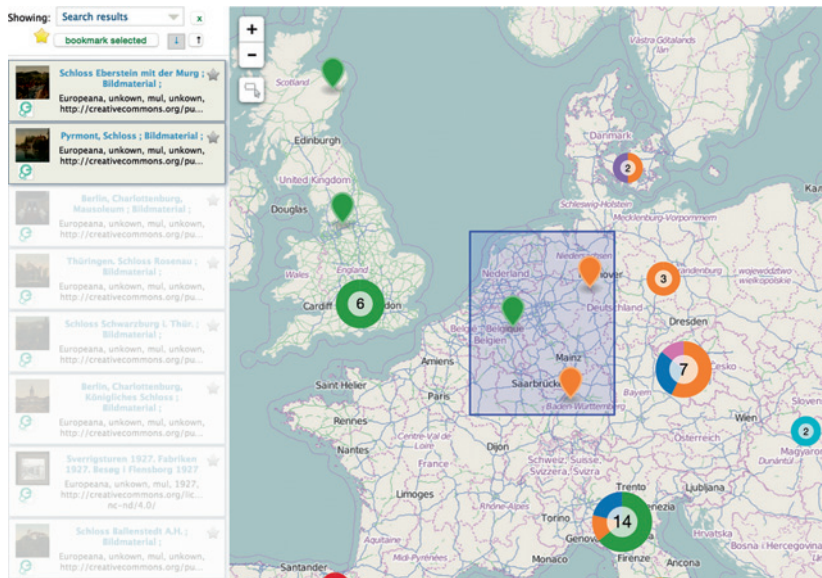


Abb. 3: Beispiel einer Geo-Visualisierung in Kombination mit Metadaten-Statistiken.

ren. Es erfolgt somit eine dynamische Verlinkung von Web-Inhalten mit miteinander in Beziehung gesetzten Long-Tail-Inhalten.

3.3 Identifikation und Bereitstellung der Inhalte

Eine weitere Kernfrage, die im Projekt beantwortet werden muss, ist, wie sich Ergebnisse von verschiedenen Quellen sinnvoll aggregieren lassen. Zwei Aspekte spielen dabei eine wichtige Rolle: die Harmonisierung der Metadaten und die intelligente Reihung der Ergebnisse aus verschiedenen Quellen.

Als Metadatenharmonisierung wird eine Abbildung von heterogenen Datenquellen auf ein einheitliches Format und Vokabular bezeichnet. Ziele sind die Gewährleistung der Interoperabilität und Integrationsfähigkeit der Daten. Das EEXCESS-Metadatenmodell (Orgel et al. 2015) erweitert das Europeana-Datenmodell (EDM⁵) mit Provenienz-Informationen, wobei die W3C PROV-Ontologie⁶ zum Einsatz kommt. Dabei muss einmalig die entsprechende Abbildung vom Quelldatenmodell auf das EEXCESS-Metadatenmodell definiert werden. Die tatsächliche Transformation der Ergebnisdaten erfolgt während der Auslieferung der Daten. Das hat den Vorteil, dass die Daten nicht im EEXCESS-Format vorgehalten werden müssen und somit keine Speicherung der Daten auf Drittsystemen und kein Datenänderungsmanagement nötig ist.

⁵ <http://pro.europeana.eu/edm-documentation>

⁶ <http://www.w3.org/TR/prov-o/>

Aufbauend auf den so harmonisierten Daten implementiert EEXCESS ein inhaltsbasiertes, föderiertes Empfehlungssystem (Ricci et al. 2011, Kern et al. 2014). Die Aufgabe von Empfehlungssystemen ist die Unterstützung von Nutzern und Nutzerinnen, denen entweder die Erfahrung oder die Kompetenz zur Erfassung und Bewertung einer großen Anzahl potenzieller Ressourcen fehlt (Resnick and Varian 1997). Empfehlungssysteme können somit als personalisierte und kontextualisierte Filter für Inhalte angesehen werden.

Die Herausforderung einer solchen Föderation ist die Selektion geeigneter Quellen auf Basis des zuvor ermittelten Kontexts, die Umformulierung der Suchanfragen in das entsprechende Quellenvokabular, sowie die intelligente Aggregation der Ergebnisse der Quellen. Durch intelligente Selektion der Quellen kann z. B. garantiert werden, dass für den Anwendungskontext „Schreiben eines wissenschaftlichen Artikels“ nur wissenschaftliche Publikationen vorgeschlagen werden.

3.4 Präsentation der Ergebnisse

Die Qualität von vorgeschlagenen Inhalten hängt nicht nur von der Auswahl der Inhalte ab, sondern auch von deren Präsentation und Darstellung (Herlocker et al. 2004, Shani und Gunawardana 2011). Visualisierungen spielen dabei eine wichtige Rolle. Vor allem im Umgang mit der Informationsüberflutung nutzen Visualisierungen den hocheffizienten menschlichen Wahrnehmungsapparat zur Verdeutlichung von Zusammenhängen (Shneiderman 1996). Visualisierungen sind somit auch für die Dar-



Abb. 4: Beispiel einer generischen Visualisierung für interaktives, facettiertes Browsing.

stellung von kontextualisierten Long-Tail-Ergebnissen geeignet, um diese zu explorieren und Zusammenhänge zu entdecken (Swearingen und Shina 2001).

Die bereits diskutierte Diversität von Long-Tail-Inhalten stellt dabei spezielle Anforderungen an Visualisierungen. Des Weiteren sind ohne Vorwissen verständliche, bekannte Visualisierungskonzepte, wie kartenbasierte Darstellungen oder Zeitstrahlen, vorzuziehen, um die Einstiegshürde niedrig zu halten.

In EEXCESS wird zudem die Art und Auswahl der Visualisierungen durch das verwendete Datenformat eingeschränkt, welches minimal Metadaten wie Titel, Autor/Autorin, Beschreibung, und eine geo-spatiale sowie temporale Komponente enthält. EEXCESS kombiniert dabei gängige Metaphern, wie z. B. eine Kartendarstellung und eine Zeitleiste, mit Statistiken über Metadaten (Sabol et al. 2014). Abbildung 3 zeigt ein Beispiel einer solchen Kartendarstellung für gefundene Long-Tail-Inhalte.

Alternative Visualisierungskonzepte sind unabhängig vom speziellen Metadatenentypen und behandeln alle Metadaten gleichwertig als textuelle Inhalte. Damit können zwar beliebige Metadaten visualisiert werden, jedoch ist die Visualisierung im Vergleich weniger aussagekräftig. Ein Beispiel einer solchen Visualisierung ist die FacetScape (Seifert, Jurgovsky und Granitzer 2014), in der die Metadatenattribute als Voronoizellen und die entsprechenden Metadatenattribute mit Hilfe einer Tag Cloud dargestellt werden (s. Abb. 4). Die FacetScape ermöglicht die interaktive Filterung gefundener Ressourcen entlang

von Metadaten, das sogenannte facettierte Browsing, und somit eine effiziente Navigation empfohlener Inhalte.

Aufgrund der Wichtigkeit solcher interaktiver Navigationswerkzeuge speziell für Long-Tail-Inhalte, konzentriert sich EEXCESS weiterhin auf die Entwicklung neuer Visualisierungen, welche auch außerhalb des gesamten EEXCESS-Frameworks nutzbar sind.

4 Resümee

Wissenschaftlich-kulturelle Long-Tail-Inhalte stellen eine wertvolle Informationsquelle dar, welche leider im heutigen Web einen zu geringen Stellenwert einnimmt. Verantwortlich dafür sind fehlende Mechanismen, Long-Tail-Inhalte Nutzern und Nutzerinnen in geeigneter Art und Weise zu Verfügung zu stellen. Die heute gängigen Strategien der Suchmaschinen-Optimierung und des Social Media-Marketings eignen sich dafür nur bedingt.

Wichtig ist die Kontextualisierung und automatisierte Bereitstellung der Inhalte, damit diese ihren vollen Mehrwert entfalten können. EEXCESS nimmt sich dieser Aufgabe an und entwickelt entsprechende Technologien, welche als Open-Source-Lösungen zu Verfügung stehen. Der Quellcode für die verschiedenen Komponenten ist im EEXCESS GitHub Repository verfügbar⁷. Die Browser Ex-

⁷ <https://github.com/EEXCESS/eexcess>

ension kann vom Chrome Webstore in den Chrome Browser installiert werden⁸. Wir hoffen, damit einen ersten Schritt in neue Nutzungsmöglichkeiten für offene wissenschaftliche und kulturelle Inhalte zu schaffen und die in diesen Inhalten vorhandenen Wissensschätze einer breiten Nutzung zuzuführen.

Literatur

- Chris Anderson (2004). The Long Tail. *Wired* 12, 10 (October 2004). http://www.wired.com/wired/archive/12.10/tail_pr.html [März 2015].
- Albert-Laslo Barabasi, Reka Albert and Hawoong Jeong (2000). Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications* 281, 1–4 (2000), 69–77.
- Sandra Carberry, Stephan Weibelzahl, Alessandro Micarelli and Giovanni Semeraro (Eds.) (2013). *User Modeling, Adaptation, and Personalization – 21th International Conference, UMAP 2013, Rome, Italy, June 10-14, 2013, Proceedings*. Lecture Notes in Computer Science, Vol. 7899. Springer.
- Gerhard Fischer (2001). User Modeling in Human-Computer Interaction. *User Modeling and User-Adapted Interaction* 11, 1–2 (3 2001), 65–86. DOI:<http://dx.doi.org/10.1023/A:1011145532042>
- Michael Granitzer, Christin Seifert, Silvia Russegger and Klaus Tochtermann (2013). Unfolding Cultural, Educational and Scientific Long-Tail Content in the Web. In *UMAP Extended Proceedings*, Shlomo Berkovsky, Eelco Herder, Pasquale Lops and Olga C. Santos (Eds.), Vol. 997. <http://ceur-ws.org/Vol-997/umap2013project1.pdf>.
- Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen and John T. Riedl (2004). Evaluating Collaborative Filtering Recommender Systems. *ACM Trans. Inf. Syst.* 22, 1 (Jan. 2004), 5–53. DOI: <http://dx.doi.org/10.1145/963770.963772>.
- Roman Kern, Kris Jack and Michael Granitzer (2014). *Recommending Scientific Literature: Comparing Use-Cases and Algorithms*. Technical Report. <http://arxiv.org/abs/1409.1357>.
- Thomas Orgel, Martin Höffernig, Werner Bailer and Silvia Russegger (2015). A Metadata Model and Mapping Approach for Facilitating Access to Heterogeneous Cultural Heritage Assets. *International Journal on Digital Libraries*. Semantic Digital Archives Special Issue.
- Paul Resnick and Hal R. Varian (1997). Recommender Systems. *Commun. ACM* 40, 3 (March 1997), 56–58. DOI: <http://dx.doi.org/10.1145/245108.245121>.
- Francesco Ricci, Lior Rokach, Bracha Shapira and Paul B. Kantor (Eds.) (2011). *Recommender Systems Handbook*. Springer.
- Vedran Sabol, Gerwald Tschinkel, Eduardo Veas, Patrick Hoefler, Belgin Mutlu and Michael Granitzer (2014). Discovery and Visual Analysis of Linked Data for Humans. In *The Semantic Web Conference – ISWC 2014, Proceedings Peter Mika et al. (Eds.)*.
- Jörg Schlötterer, Christin Seifert and Michael Granitzer (2014). Web-based Just-In-Time Retrieval for Cultural Content. In *PATCH '14: Proceedings of the 7th International ACM Workshop on Personalized Access to Cultural Heritage*.
- Christin Seifert, Johannes Jurgovsky and Michael Granitzer (2014). FacetScope: A Visualization for Exploring the Search Space. In *Proc. International Conference on Information Visualization (IV)*. IEEE Computer Society, 94–101. DOI: <http://dx.doi.org/10.1109/IV.2014.49>.
- Christin Seifert, Jörg Schlötterer and Michael Granitzer (2015). Towards a Feature-Rich Data Set for Personalized Access to Long-Tail Content. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*. ACM, New York, NY, USA.
- Guy Shani and Asela Gunawardana. 2011. Evaluating Recommendation Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Eds.). Springer US, 257–297. DOI: http://dx.doi.org/10.1007/978-0-387-85820-3_8.
- Ben Shneiderman. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *IEEE Visual Languages*. College Park, Maryland 20742, U.S Shneiderman A., 336–343.
- Kirsten Swearingen and Rashmi Sinha. 2001. Beyond Algorithms: An HCI Perspective on Recommender Systems. In *ACM SIGIR. Workshop on Recommender Systems*, Vol. 13, Numbers 5–6. 393–408.



Dr. Christin Seifert
Universität Passau
Lehrstuhl für Medieninformatik
Innstraße 33a
94032 Passau
christin.seifert@uni-passau.de
<http://www.uni-passau.de/medieninformatik/>

Dr. Christin Seifert ist Post-Doc an der Professur für Medieninformatik an der Universität Passau und arbeitet an den EU-Projekten CODE und EEXCESS mit. Sie promovierte im Themenbereich Interaktives Maschinelles Lernen an der Technischen Universität Graz und erhielt ihr Diplom von der Universität Chemnitz für ihre Arbeit im Bereich Künstliche Intelligenz. Sie war am Know-Center und am Joanneum Research in Graz an mehreren nationalen und EU-geförderten Forschungsprojekten beteiligt. Christin Seifert publizierte über 70 begutachtete Publikationen in den Forschungsbereichen Maschinelles Lernen, Informationsvisualisierung und Bilderkennung.



Prof. Dr. Michael Granitzer
Universität Passau
Lehrstuhl für Medieninformatik
Innstraße 33a
94032 Passau
michael.granitzer@uni-passau.de
<http://www.uni-passau.de/medieninformatik/>

Prof. Dr. Michael Granitzer ist Professor für Medieninformatik an der Universität Passau. Er wurde an der Technischen Universität Graz im Fachgebiet Telematik promoviert und war wissenschaftlicher Leiter der außeruniversitären Forschungseinrichtung Know-Center in Graz. Er leitete mehrere national und international geförderte Forschungsprojekte, wie z. B. das EU Projekt CODE oder das integrierte Projekt EEXCESS. Michael Granitzer publizierte über 150 begutachtete Publikationen in den Forschungsschwerpunkten Text und Data Mining, Linked Open Data, Visual Analytics und Information Retrieval.

⁸ purl.org/eexcess/clients/chrome-extension