

Web@rchiv Österreich

Überblick und Aktivitäten zur Flüchtlingskrise

Im Jahr 2008 wurde von der Österreichischen Nationalbibliothek mit dem *Web@rchiv Österreich* ein nationales Webarchiv begründet, das seit 1. März 2009 durch das österreichische Mediengesetz zur Sammlung und Präsentation von online Inhalten ermächtigt ist. In ihrer Vision 2025 sieht die Österreichische Nationalbibliothek in Zukunft den „Schwerpunkt unserer Sammlungspolitik im Online-Bereich, weil der Großteil der österreichischen Publikationen digital erscheint“ und will bei der Sammlung „usergenerierte Inhalte, öffentliche soziale Netzwerke, Blogs und neu entstandene Formate berücksichtigen, die wichtige Einblicke in Kultur und Gesellschaft des 21. Jahrhunderts ermöglichen“.

Sammelaktivitäten

Bisher wurden durch die Kombination von vier Domain Crawls, laufenden selektiven und punktuellen Event Crawls folgende Datenmengen vom *Web@rchiv Österreich* gesammelt und archiviert:

- Datenvolumen gesamt: 85 Terabyte
- Anzahl Einzeldateien: 2,7 Milliarden
- Unterschiedliche Seeds¹: 1,7 Millionen

Der Großteil der Domains entfällt dabei auf die .at Domäne inklusive der Subdomains .ac.at (Wissenschaft, Bildung) und gv.at (Verwaltung, Behörden), und inkludiert seit kurzem auch die neue Top Level Domain .wien. Der Prozess des Sammelns, das sogenannte Harvesting oder Crawling, wird inhouse durchgeführt, die physische Speicherung der Daten erfolgt im Bundesrechenzentrum.

Aufgrund der zweijährigen Intervalle zwischen einzelnen Domain Crawls und dem damit verbundenen Risiko, dass in der Zwischenzeit zahlreiche Webseiten verloren gehen, werden für ausgewählte Seiten, die häufigen Änderungen unterliegen oder thematisch von besonderem Interesse sind, in kürzeren Abständen selektive Harvestings durchgeführt. Um das aktuelle Tagesgeschehen und das politische Leben in Österreich zu dokumentieren, wurden spezielle

¹ Ein Seed ist eine Internetadresse, die der Webcrawler als Ausgangspunkt verwendet, z.B. <http://www.onb.ac.at>.

Webseiten ausgewählt, die täglich im Rahmen einer Medien- und einer Politikkollektion gesammelt werden.

Webseiten zu besonderen Anlässen und Großereignissen sind außerordentlich gefährdet (z.B. <http://euwahl09.orf.at>) und stehen meist nur für den Zeitraum des Ereignisses zur Verfügung. Die Intervalle für diese sogenannten Event Crawls werden individuell definiert. So sind bereits zu zahlreichen wichtigen Ereignissen Webinhalte mit Österreich-Bezug im *Web@rchiv Österreich* zu finden (z.B. Wahlen, Sportereignisse, Gedenkjahr Erster Weltkrieg, Eurovision Song Contest etc.).

Flüchtlingskrise

Nachdem bedingt durch internationale Entwicklungen im Sommer 2015 ein massiver Anstieg der Flüchtlingszahlen eingetreten ist, wurde rasch klar, dass es sich um Ereignisse von nationaler Bedeutung handelt, die auch im *Web@rchiv Österreich* dokumentiert werden sollten. Zahlreiche Initiativen organisierten sich im Web, Hilfsorganisationen, Gemeinden und andere öffentliche Stellen publizierten online Inhalte zur Flüchtlingsthematik.

Seit September 2015 archiviert die Österreichische Nationalbibliothek von KuratorInnen ausgewählte Webseiten und Social Media Profile zur Flüchtlingskrise (51 Domains mit nahezu 100 Seeds). Im Rahmen der wöchentlich stattfindenden Crawls wurden bisher 215 GB an Daten gesammelt. Die Thematik wird zusätzlich über die laufende Medien- und Politikkollektion (tägliche Speicherung) abgedeckt.

Zugriff auf archivierte Webseiten

Der Zugriff auf das *Web@rchiv Österreich* ist durch das Mediengesetz geregelt und kann vor Ort in der Österreichischen Nationalbibliothek, bei ausgewählten berechtigten Bibliotheken – nicht aber online – erfolgen. Die elektronische Weiterverarbeitung der archivierten Daten ist nicht gestattet, ein Ausdruck der Webseiten kann bei Bedarf erstellt werden.

Weitere Informationen und Kontakt

<http://www.onb.ac.at/about/webarchivierung.htm>
webarchiv@onb.ac.at

In Kürze Online-Suche unter <https://webarchiv.onb.ac.at/>