



SCHWEIZERISCHE NATIONALBIBLIOTHEK NB



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Kultur BAK
Schweizerische Nationalbibliothek NB

Maya Bangerter

bei der Schweizerischen NB verantwortlich für die Suchapplikation e-Helvetica Access

„Wir hatten sehr spezifische Anforderungen, auf die SHI sehr genau einging, statt einfach standardisierte Module und Lösungen zu implementieren. Vor allem war die Datenmenge für uns eine große Herausforderung, die wir nun gemeistert haben. Die beschleunigte Indexierung und das umfassende Highlighting in den Suchergebnissen sind für uns Ergebnisse, auf die wir besonders stolz sind.“

SCHWEIZERISCHE NATIONALBIBLIOTHEK MEISTERT BIG DATA MIT OPEN SOURCE

Mit einer neuen Suchapplikation finden Nutzer schnell und komfortabel alle Treffer

Über die Schweizerische Nationalbibliothek

Bei der Schweizerischen Nationalbibliothek (NB) finden die Nutzer umfassende Informationen über die Schweiz – unabhängig davon, auf welchem Medium sie gespeichert sind. Über das Onlineportal e-Helvetica Access (www.e-helvetica.nb.admin.ch) können sie in den digitalen Publikationen und Web-Archiven der NB recherchieren. Per Volltextsuche erhalten sie dort zu ihrem Suchbegriff eine Ergebnisliste aller Inhalte.

Die Sammlung der digitalen Bibliothek e-Helvetica enthält rund 140.000 digitale Bücher, Zeitschriften und Dissertationen sowie als Besonderheit das Webarchiv Schweiz, das historische Webseiten mit Bezug zur Schweiz archiviert. Alle Objekte umfassen derzeit insgesamt Daten in der Größenordnung von circa 30 Terabyte, wobei das Webarchiv den Hauptanteil der Datenmenge ausmacht.

AUF EINEN BLICK:

- Branche: Öffentlicher Dienst - Bibliothek
- Ziele: Schnelle vollständige Trefferlisten bei Suchabfragen; bessere Usability und Convenience für die Nutzer
- Lösungen: On-Site Search
- Technologien: u.a. Apache Solr, Apache Solr Cloud, Apache NiFi zur Integration des WARC-Indexer

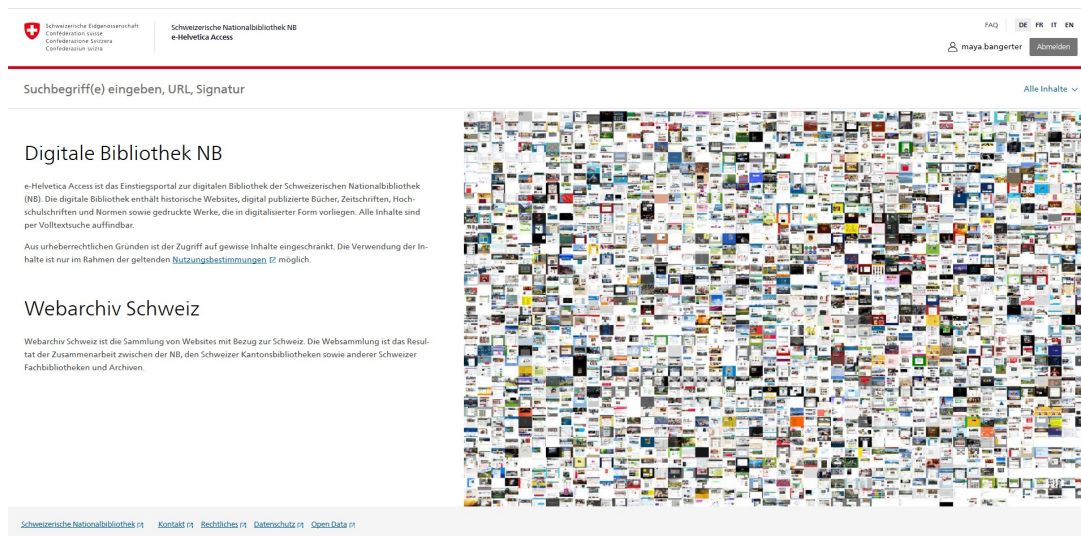


Abb 1: das Onlineportal e-Helvetica Access der Schweizerischen NB

Herausforderung

Die Datenmenge der digitalen Bibliothek nahm immer weiter zu. Doch bereits die vorhandenen Daten waren nicht vollständig im Solr Index gespeichert. Die Daten der archivierten Webseiten befanden sich in einer separaten Datenbank. Die vorhandenen IT-Systeme waren nicht ausreichend dimensioniert, um das Webarchiv mit angemessenem Aufwand vollständig in die Volltextsuche zu integrieren.

Nach der Eingabe eines Suchbegriffs mussten die Nutzer daher relativ lange warten, bis erste Ergebnisse angezeigt wurden. Um den Nutzern eine schnellere und umfassende Recherche zu ermöglichen, sollte die Suchfunktion erneuert werden. Als die gesamten IT-Systeme der Schweizerischen NB auf die Infrastruktur der Schweizerischen Nationalphonothek migriert werden sollten, war das ein guter Anlass.

Machbarkeitsstudie und Zielsetzung

Um die erforderlichen Maßnahmen zu planen, startete die Schweizerische NB das Projekt mit einer Machbarkeitsstudie. Sie beauftragte damit die SHI, die nach einem gemeinsamen Workshop im Juni 2017 unter anderem folgende Ergebnisse festhielt:

- **Neuaufbau des Volltextindex**
Ein Volltextindex, der auch die großen Datenmengen des Webarchivs berücksichtigt, ist mit Hilfe des Open Source Tools Warc-Indexer und Apache NiFi in einem angemessenen Zeitraum realisierbar.
- **Indexierung der Metadaten**
Die Metadaten bieten wichtige Informationen zu den einzelnen Objekten, die bei einer Suche ebenfalls berücksichtigt werden sollten. Die Angaben zu Autor, Publikationstyp, Verlag und anderen Aspekten sind in XML Dateien gespeichert. Für die Indexierung benötigt die Nationalbibliothek eine maßgeschneiderte Lösung.

Die Zielsetzung war, dass die Suchfunktion der digitalen Bibliothek leistungsfähiger und den modernen Anforderungen an die Usability gerecht werden sollte. Dazu zählen nicht nur Geschwindigkeit und Vollständigkeit, sondern beispielsweise auch Funktionen wie Facettierung und Highlighting in den Trefferlisten. Die Suche sollte in den vier Sprachen Deutsch, Englisch, Französisch und Italienisch zur Verfügung stehen.

Lösung

Die Schweizerische NB beauftragte 2018 die Experten der SHI. In enger Zusammenarbeit setzten sie gemeinsam folgende Maßnahmen um:

- **Verteilter Index - Apache Solr Cloud:**
Basierend auf den vorhandenen Daten schätzten die Solr Experten der SHI das erforderliche Sizing ab und installierten auf einer Linux-Plattform die zu diesem Zeitpunkt neueste Solr Cloud Version.
- **Indexierung mit Apache NiFi:**
Sie realisierten die Indexierung aller digitalen Objekte, so dass mit einer Anfrage sowohl die Non-Web- als auch die Web-Daten durchsucht werden. Die Objekte des Webarchivs konnten sie mit Hilfe der Open-Source -Komponente Warc-Indexer und Apache NiFi in weniger als sechs Wochen indexieren. Für die Indexierung der Metadaten entwickelten die Schweizerische NB und SHI gemeinsam eine neue, maßgeschneiderte Lösung.

Die neue Suchanwendung wurde unabhängig von dem vorherigen System aufgebaut und löste dieses komplett ab.

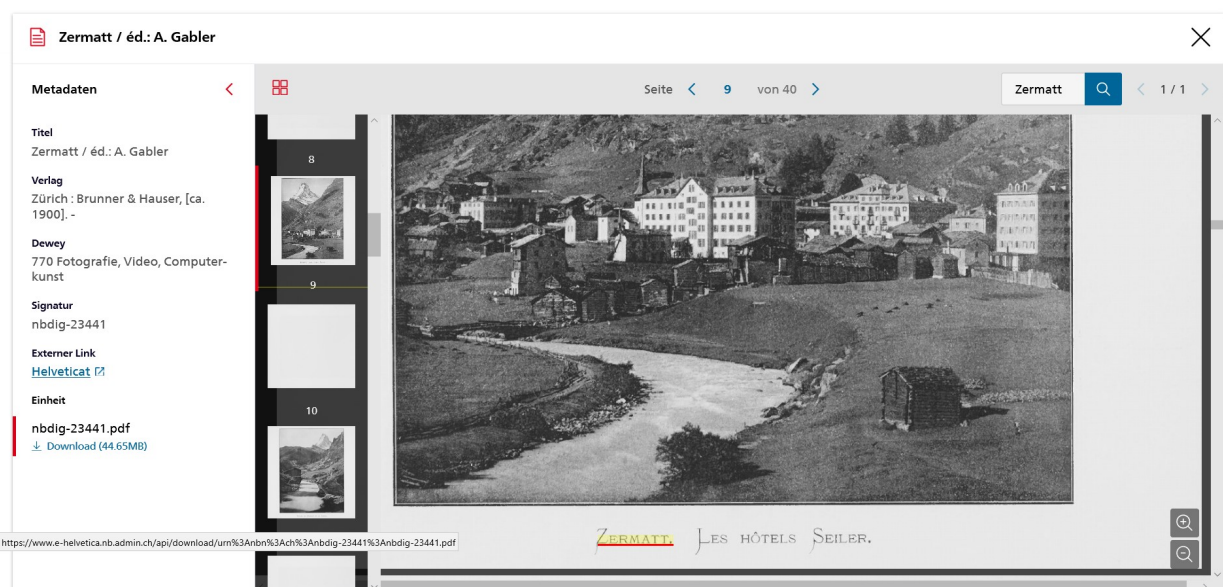


Abb 2: Beispiel für die farbliche Markierung des Suchbegriffs in einem Dokument

Ergebnis

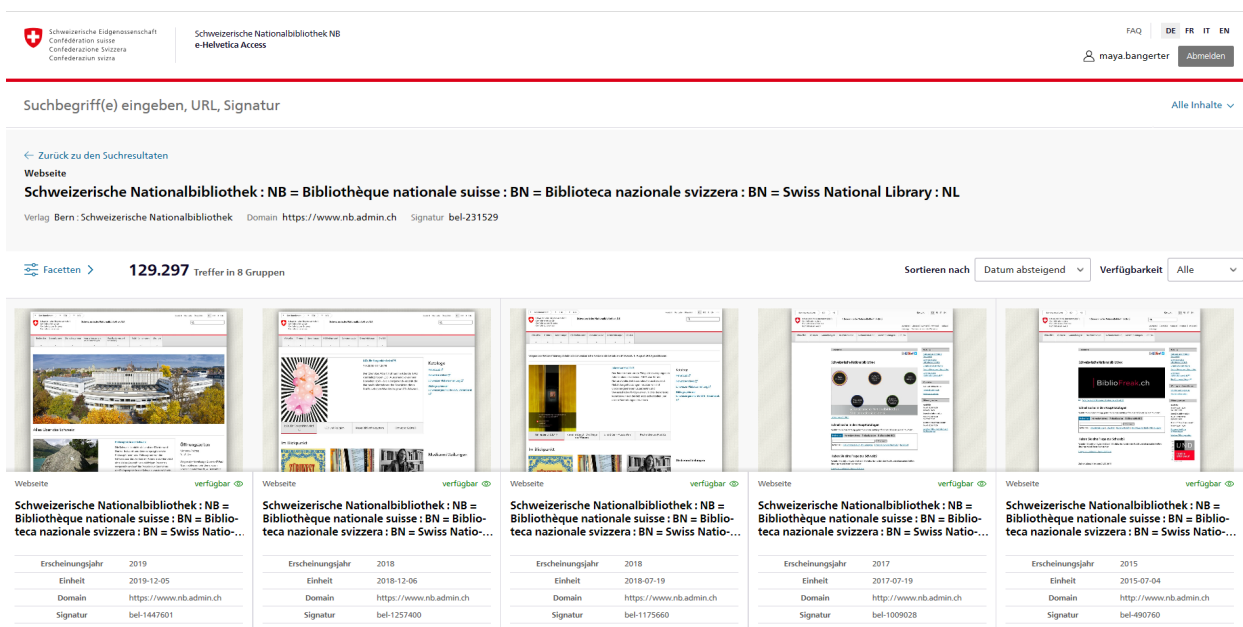
Seit Juli 2019 steht den Nutzern über e-Helvetica Access eine leistungsfähige und moderne Suchapplikation zur Verfügung. Sie können in vier Sprachen nach selbst gewählten Begriffen suchen und erhalten bereits während der Eingabe potenzielle Treffer angezeigt.

Die hohe Geschwindigkeit, mit der alle digitalen Objekte aufgelistet werden, verbessert die Nutzung deutlich. Dies wird auch durch eine optimale Abstimmung zwischen dem Solr System und dem neuen Frontend erreicht, welches gleichzeitig durch einen anderen Dienstleister entwickelt wurde.

Eine komfortable Suche und hervorragende Usability von e-Helvetica Access ist das Ergebnis der engen Zusammenarbeit zwischen den Teams der Schweizerischen NB, der Frontend-Entwickler und der SHI. Beispielsweise ermöglicht die Funktion des Highlighting, dass die Nutzer einen schnellen Überblick über die Suchergebnisse erhalten, die für sie relevant sind.

Bereits die Vorschau der Suchergebnisse zeigt bei verfügbaren Objekten den farblich markierten Suchbegriff in einem Snippet an. Der Nutzer kann dann mit einem Klick das gesamte Dokument öffnen und beispielsweise das digitalisierte Buch komplett einsehen, sofern die Urheberrechte dies zulassen. Im gesamten Dokument ist der Suchbegriff ebenfalls an allen Stellen farblich markiert. Besonders hilfreich ist die Spalte, die einen Überblick über das Buch bietet. Dank eines farbigen Strichs erkennt der Nutzer dort auf den ersten Blick, auf welcher Seite sich an welcher Position der Suchbegriff befindet.

Alle Interessenten erhalten online einen vollständigen Überblick über die umfassende digitale Bibliothek der Schweizerischen NB. In zahlreichen Dokumenten können sie ohne Anmeldung oder andere Einschränkungen online recherchieren. Bei urheberrechtlich geschützten Objekten erfahren sie, an welchem Standort der Nationalbibliothek sie das Dokument einsehen können.



The screenshot shows the search results page for the Swiss National Library (NB). The search term is "Schweizerische Nationalbibliothek : NB = Bibliothèque nationale suisse : BN = Biblioteca nazionale svizzera : BN = Swiss National Library : NL". The results are sorted by date in descending order, showing 129,297 hits in 8 groups. The first five results are displayed as cards, each showing a thumbnail of the document cover and a snippet of the text. The snippet highlights the search term in red. Below each card, there is a table with the following information:

Errscheinungsjahr	Einheit	Domain	Signatur
2019	2019-12-05	https://www.nb.admin.ch	bel-1447601
2018	2018-12-06	https://www.nb.admin.ch	bel-1257400
2018	2018-07-19	https://www.nb.admin.ch	bel-1175660
2017	2017-07-19	http://www.nb.admin.ch	bel-1009028
2015	2015-07-04	http://www.nb.admin.ch	bel-490760

Abb 3: Vorschau der Suchergebnisse