

True Online Publishing: Im Internet publizierte Information soll referenzierbar und verfügbar bleiben

Hella Dietz

(factline Krisper Fabro Harnoncourt OEG
hella.dietz@factline.com)

1. Einleitung

Die besonderen Möglichkeiten, die das Internet bietet, sind geläufig: Während in der Welt des Print eine starke Vorselektion vorgenommen wird, kann im World Wide Web jeder Nutzer veröffentlichen, was ihm interessant erscheint. Information kann schnell und kostengünstig verbreitet werden, Hyperlinks ermöglichen die Aufweichung der linearen Informationsvermittlung (Dokumente können von vorne nach hinten oder den jeweiligen Querverweisen folgend gelesen werden) und ermöglichen es, auf andere Dokumente zu verweisen. Einige dieser Möglichkeiten können jedoch nur eingeschränkt genutzt werden: Zwar können alle Nutzer Information veröffentlichen, Quelle und Plausibilität der Information sind oft nur schwer nachvollziehbar. Zwar kann mittels URL auf jede beliebige Seite im WWW gelinkt werden, ob und wie lange das Dokument dort *in unveränderter Form* abgelegt ist, ist unsicher. So bleibt den Nutzern meist keine andere Wahl, als es abzuspeichern oder auszudrucken. Die Folge ist Vervielfältigung und Redundanz von Information.

Problematisch ist somit die herkömmliche Art des Verlinkens und, wichtiger noch, der Umgang mit ins Internet gestellter Information. Im nächsten Abschnitt sollen diese beiden Probleme kurz skizziert und existierende Lösungsansätze vorgestellt werden. Vor allem für das zweite Problem gibt es zwar nationale Lösungen und bereichsspezifische Ansätze, aber unseres Wissens noch keine übergreifende, handhabbare Lösung. Das von factline entwickelte Konzept des "True Online Publishing" soll sich zu einer solchen Lösung entwickeln. In Abschnitt 3 erläutere ich, was bereits umgesetzt ist und welche Lösungsvorschläge wir für noch ungelöste Problemfelder angedacht haben. Diese möchte ich im Anschluß zur Diskussion stellen.

2. Probleme des Verfügbarmachens von Information – Status Quo

2.1 Das Problem, Information verfügbar zu halten (Archivierung):

In letzter Zeit gibt es verstärkte Bestrebungen, Inhalte des Internet zu archivieren.

Exemplarisch: "Es steht heute außer Zweifel, dass der Bereich der elektronischen Medien in den letzten zehn Jahren zu einem unverzichtbaren Teil der modernen Informationsvermittlung und damit der kulturellen Identität eines Landes geworden ist. Will Österreich nicht diesen immer wichtiger werdenden Anteil seines geistigen Kulturerbes auf längere Sicht hin verlieren, muss dringend Vorsorge für eine Langzeitarchivierung der in Österreich publizierten elektronischen Medien getroffen werden."

(Schmidt 2000) Über dieses Thema wurde auch auf der im Januar veranstalteten

Konferenz "cultH2 – die Zukunft des /digitalen/ kulturellen Erbes" diskutiert (siehe http://www.culth.org/d0_index.htm). Außerdem gibt es verschiedene Projekte zur Publikation und Archivierung wissenschaftlicher Publikation im Netz, bspw. das Projekt "Dissonline" der Deutschen Bibliothek (siehe: <http://www.dissonline.de>). Diese Ansätze gehen entweder davon aus, daß der Autor/die Autorin die Information bereits selbst für andere verfügbar halten möchte oder ein unabhängiges Institut nach bestimmten Kriterien und gewissermaßen "auf Verdacht" archiviert, was als archivierungswürdig angesehen wird. Es fehlt allerdings ein ergänzendes Konzept, das es Internetnutzern ermöglicht, zuverlässig auf ein im World Wide Web publiziertes Dokument anderer Autoren zu verweisen, wenn dieser nicht schon selbst entschieden hat, das Dokument verfügbar zu halten und es auch (noch) nicht anderweitig archiviert ist.

2.2 Das Problem des Verweises mittels URL:

Herkömmlicherweise wird auf ein Dokument verwiesen, indem der "Uniform Resource Locator" (URL) der Seite angegeben wird – nicht das Dokument selbst wird identifiziert, sondern die Seite (der Ort), an dem es gegenwärtig abgelegt ist: "In actuality, the URL is simply an address masquerading as an identifier, and relying on it to identify a unique digital resource is analogous to using a home address in place of a social security number." (Payette 1998)

Die Notwendigkeit von "unbreakable links" wurde genau genommen schon vor Entstehung des Internets, das wir heute kennen, betont: Ted Nelson entwickelt seit 1965 XANADU, ein alternatives Konzept für ein Netz von Hypertexteinheiten. Nelsons Visionen sind für die einen Fundament der Hypertextbegeisterung (Kuhlen 1991:69), für andere ist Xanadu "the longest-running vaporware story in the history of the computer industry" (Wolf 1995). Es gibt jedoch zwei weitere Ansätze, das Problem oder zumindest Teilaspekte zu lösen:

1. Eine Arbeitsgruppe der "Internet Engineering Task Force" (IETF) arbeitet seit geraumer Zeit an der Erweiterung des URL-Schemas. Jedes Dokument erhält – analog zur ISBN – eine eindeutige Bezeichnung, den "Uniform Resource Name" (URN). In einer Datenbank werden URN und zugehöriger URL abgelegt. Ändert sich der URL, wird dies in der Datenbank vermerkt. Zusätzliche Information wie Format, Quelle, Größe etc. wird in den sog. "Uniform Resource Characteristics" (URC) gespeichert (siehe <http://www.ietf.org/html.charters/urn-charter.html>). Noch gibt es keine Einigung über die Einführung des Standards, wenngleich einzelne Projekte auf dem URN-Konzept basieren: "Persistent URL" (<http://purl.oclc.org/oclc/purl/sightings>) und die Initiative "Digital Object Identifier" (DOI). Auf letztere möchte ich kurz eingehen: Von der "Association of American Publishers" (AAP) konzipiert, erhalten Online-Publikationen eine eindeutig identifizierbare Nummer, die von einem Resolver System ("Handle System", siehe: <http://www.handle.net/>) auf eine (oder mehrere mögliche) URL umgeleitet wird. Der jeweils Berechtigte kann die URLs bei Bedarf ändern. Allerdings ist die Verwendung von DOIs (noch) auf Verlage beschränkt und kommt aus Kostengründen (US\$ 1000 pro DOI) nicht für eine breitere Anwendung in Frage (siehe <http://www.doi.org>).

Diese Standardisierungsbestrebungen lösen das Problem der "broken links", nicht jedoch das der Verfügbarkeit: Wenn ein Dokument aus dem Netz genommen wird, kann nicht mehr darauf zugegriffen werden. Alternative Lösungen sollten jedoch in jedem Fall mit URN kompatibel (als offener Standard) konzipiert werden.

2. Neben diesen institutionalisierten Standardisierungsbestrebungen haben auch Anbieter von Content Management Systemen Lösungsansätze implementiert. In datenbankbasierten Systemen wird die eindeutige Identifikationsnummer der Datenbank genutzt, um ortsunabhängig zu verweisen. Das in anderem Zusammenhang von Hyperwave entwickelte Konzept bidirektionaler Links erreicht, daß der Zitierende zumindest informiert wird, wenn die zitierte Information von der Seite entfernt wird. Allerdings kann auch dieses Konzept nicht garantieren, daß sich das Dokument nicht zwischenzeitlich ändert und löst das Problem der Archivierung nicht.

3. factlines Konzept des "True Online Publishing"

Im folgenden sollen bereits implementierte Aspekte vorgestellt werden und Konzepte zur Weiterentwicklung präsentiert und zur Diskussion gestellt werden.

3.1 TOP – Status Quo:

factline entwickelt und vertreibt eine Software für Wissensgemeinschaften – den factline Community Server (FCS). Um Wissensgemeinschaften zu ermöglichen, innerhalb von auf dem FCS laufenden Plattformen zuverlässig auf interessante Dokumente anderer Autoren zurückgreifen zu können, hat factline "True Online Publishing" entwickelt: factline garantiert Speicherung und dauerhafte Verfügbarkeit der Information unter einer eindeutigen Identifikationsnummer. Jeder publizierten Informationseinheit kann eine eindeutige und unveränderbare "factID" zugeordnet werden. Unter Angabe der factID kann die Informationseinheit zitiert bzw. mit anderen Informationen verknüpft werden. Jede factID erhält zudem ein Suffix zur Versionskontrolle: Wird ein Text verbessert und aktualisiert, versieht der FCS das Dokument mit einer höheren Versionsnummer. Der Verweis kann sich dann wahlweise auf eine bestimmte Version beziehen (wenn beispielsweise auf Fehler eines Dokumentes hingewiesen werden soll) oder auf die aktuellste verfügbare Version (Wettervorhersagen, Wechselkurse, Preislisten u.ä.).

Da jede (publizierte) Änderung des Dokuments eine höhere Versionsnummer zur Folge hat, kann die Information nicht unbemerkt verändert werden. Die Speicherung garantiert, daß auch ältere und nicht mehr im Internet verfügbare Versionen weiterhin unter der zugehörigen factID abrufbar sind, Verweise bleiben bestehen. Mit jeder TOP publizierte Information ist außerdem Information über Autor, Copyright und Publikationsdatum verknüpft. Dies ist ein erster Schritt, um die in Punkt 2 geschilderten Probleme zu lösen, weitere Schritte sollen folgen.

3.2 Weiterentwicklung des Konzepts

Im Moment arbeitet factline daran, dieses Konzept über den FCS hinaus nutzbar zu machen: User sollen für ein beliebiges Dokument, auf das verwiesen werden soll, eine factID beantragen können, das Dokument wird in einer Datenbank abgelegt, auf die bei Bedarf zurückgegriffen werden kann. Zu diskutierende Eckpunkte:

Archivierung:

Für die Langzeitarchivierung von mehr als 500 Jahre alten Printdokumenten verfügen Bibliotheken seit langem über taugliche Methoden, wesentlich schwieriger ist es, digitale Daten zugänglich zu halten. Voraussetzungen:

- genügend Speicherplatz (nicht kostenkritisch, solange es um reine Texte geht)
- Konvertierbarkeit von Dateiformaten (wichtig: offener Standard)
- Kompatibilität von Speichermedien

Retrieval:

Drei Möglichkeiten, auf "TOP-publizierte" Informationen zuzugreifen:

- Benutzer folgt einem Link oder stößt beim Browsen auf die Information
- Aufruf mittels der bekannten factID über ein Trustcenter (der Benutzer weiß bereits, was er eigentlich sucht)
- Aufruf über eine Suchabfrage (Benutzer sucht nach hilfreicher Information)

Der dritte Fall stellt an TOP die größte Herausforderung: Sinnvolle Suchabfragen erfordern den Einsatz von Ontologien, "intelligent indexing", Rating (Grad der Vernetzung) etc. Außerdem sollen Bewertungssysteme (wie hilfreich war die Information, wie hilfreich war die Bewertung) integriert werden.

Einspeisen von Information:

Um die unter "Retrieval" geschilderten Suchanforderungen umsetzen zu können, müssen folgende Voraussetzungen beim Einspeisen erfüllt werden:

- Information und Autor müssen authentifiziert werden (digitale Signatur)
- sinnvolle Beschlagwortung (Metadaten (Standard: Dublin Core))

Falls zudem die Nutzergruppe der Zitierenden Zielgruppe sein soll, muß es ein umsetzbares Modell geben, das die Urheberrechte der Information wahrt.

Berechtigungen:

Für online publizierte Information soll die Möglichkeit bestehen, zu definieren, welche Benutzergruppen leseberechtigt sind. Dies wird für jede Informationseinheit in den Metadaten festgehalten. Nach der Publikation kann der Benutzerkreis zwar noch erweitert, aber nicht weiter eingeschränkt werden: Dahinter steht die Überlegung, daß jede Information nach ihrer Publikation Teil des persönlichen Wissensnetzwerkes der Zitierenden ist.

Urheberrecht:

Will der Autor Information selbst verfügbar machen, können urheberrechtliche Fragen direkt geklärt werden. Weit schwieriger ist die Abspeicherung von Informationen auf Anfrage des Zitierenden, da in diesem Fall Urheber und/oder Rechteinhaber nicht in den Prozess involviert ist. Die Rechtslage ist international noch weitgehend ungeklärt, zahlreiche Anwendungsbeispiele zeigen jedoch, daß die Archivierung fremder Information bereits praktiziert wird und urheberrechtliche Fragen den Betrieb offensichtlich nur wenig stören. Beispiele:

- Google (www.google.com) erstellt und speichert "Schnappschüsse" von jeder untersuchten Seite. Der User kann direkt auf den "Cached"-Link zugreifen, wenn der Originallink nicht (mehr) verfügbar ist. Anfang Anfang 2001 macht Google zudem das von Deja.com übernommene Usenet-Archiv mit mehr als 500 Mio. Newsgroup-Postings seit 1995 zugänglich.
- Das Non-Profit-Projekt "The Internet Archive" (www.archive.org) speichert seit 1996 in regelmässigen Abständen Schnappschüsse aller öffentlich zugänglichen Websites ab (seit Ende Oktober öffentlich zugänglich).

Zu diesen Punkten möchte ich unsere Überlegungen zur Diskussion stellen.

Literatur:

Kuhlen, Rainer (1991): "Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissensbank", Berlin, Heidelberg, New York.

Payette, Sandra (1998): "Persistent Identifiers on the Digital Terrain", in: RLG DigiNews, Volume 2, gefunden im Internet unter:
<http://www.rlg.org/preserv/diginews/diginews22.html#Identifiers>, Stand: 17.01.02

Schmidt, Alfred (2000): "Einführung: Die Notwendigkeit der Archivierung elektronischer Publikationen", Nationalbibliothek, gefunden im Internet unter:
<http://www.onb.ac.at/about/projekte/arch-el-medien.htm>, Stand 17.01.02

Wolf, Gary (1995): The Curse of Xanadu, in: WIRED, Volume 3, Issue 6, gefunden im Internet unter: http://www.wired.com/wired/archive/3.06/xanadu_pr.html, Stand 17.01.02

Acknowledgements:

Das in diesem Artikel dargestellte Konzept wurde von factline unter anderem in einem Arbeitskreis zum Thema entwickelt. Die TeilnehmerInnen des Arbeitskreises: Max Harnoncourt, Lotte Krisper, Paul Meinl, Sebastian Udulutsch.