

Digitale Editionen als Web-Services

,
immanuel.normann@pagina-tuebingen.de
pagina GmbH, Deutschland

Verstehen wir unter einer digitalen Edition eine „erschließende Wiedergabe historischer Dokumente“, welche dem digitalen Paradigma folgt, indem sie die gegenwärtigen technischen Möglichkeiten berücksichtigt (cf. Sahle 2013: 138, 148), dann stellt sich die Frage, welche technischen Möglichkeiten zu welchem Zweck eingesetzt werden sollen. In diesem Beitrag wird die Überzeugung vertreten, dass digitale Editionen als zentraler Bestandteil von Forschungsumgebungen der Textwissenschaft von weit größerem Nutzen sein können, wenn sie über standardisierte semantische Web-Schnittstellen verfügen. Digitale Editionen wären dann primär als Web-Services zu verstehen, die über ihre Web-Schnittstellen mit anderen Web-Services oder mit Web-Anwendungen kommunizieren. Es wäre erst die Web-Anwendung (welche im Browser ausgeführt wird), mit der der menschliche Nutzer interagiert, wogegen alle übrige Kommunikation von Maschine zu Maschine liefere. Herkömmliche digitale Editionen sind primär auf eine Nutzung durch den Menschen allein ausgerichtet. Die im Folgenden zu begründende These ist, dass Werkzeuge der Forschungsumgebungen mit diesen herkömmlichen digitalen Editionen deshalb nur unbefriedigend ineinandergreifen, weil sie programmatisch abgeschlossen sind. Dieser Zustand ist insofern unbefriedigend, als dadurch Textforschung weit weniger vernetzt und kollaborativ vonstatten geht als dies möglich wäre.

Eine Verbesserung dieses Zustands kann natürlich nicht allein von technischen Neuerungen digitaler Editionen erhofft werden. Es sind ebenso technische Neuerungen bei allen Komponenten bestehender Forschungsumgebungen nötig (und bei Initiativen wie TextGrid auch im Gange). Dabei besteht eine wechselseitige Abhängigkeit des Entwicklungsfortschritts: Nur wenn die eine Komponente das eine neue Feature anbietet, besteht bei der anderen Komponente die Chance eines Entwicklungssprungs. Mit Blick auf diese *Koevolution* müssen also diejenigen Komponenten einer Forschungsumgebung berücksichtigt werden, die mit einer digitalen Edition im Datenaustausch stehen oder stehen sollten. Dabei ist es zielführend, sich nicht ausschließlich von der Frage leiten zu lassen, wie man digitale Editionen möglichst interoperabel zu den am weitestverbreiteten Werkzeug der Textwissenschaftler (z. B. der dominierenden Textverarbeitungssoftware) machen kann. Vielmehr sollte die Aufmerksamkeit darauf gerichtet werden, welche nützlichen Werkzeuge man schaffen

könnte, wenn man die digitalen Editionen mit bestimmten technischen Neuerungen ausstatten würde.

Im Folgenden wird daher das Umfeld digitaler Editionen innerhalb einer textwissenschaftlichen Forschungsumgebung in den Blick kommen und zwar in einer Weise, die auch noch nicht existierende Systeme mitdenkt. Dies ist möglich, wenn man eine solche Umgebung zu diesem Zweck nicht als eine Ansammlung bestehender Tools auffasst, sondern die textwissenschaftlichen Tätigkeiten identifiziert, für die man sich ohne Rücksicht auf bestehende Fertiglösungen technische Unterstützung überhaupt vorstellen kann.

Die aus informationstechnischer Sicht relevanten Tätigkeiten lassen sich in diesem Kontext sinnvoll unterteilen in: das *Lesen*, *Schreiben* und *Verwalten* von Text. Während das Lesen und Schreiben von Text in diesem Rahmen keiner weiteren Erklärung bedarf, muss näher darauf eingegangen werden, was mit Textverwaltung alles gemeint sein kann. Eine positive Definition dieses Begriffs würde wahrscheinlich keine allgemeine Zustimmung finden, daher sollen ein paar paradigmatische Beispiele zur Begriffsklärung ausreichen: Exzerpieren, Organisieren von Textschnipseln in Zettelkästen, Anlegen von Literaturlisten, Zusammenstellen eines Semesterapparats, Sortierung von Büchern, Klassifikation von Texten, Erstellen von Registern und vieles mehr – für all diese und ähnliche Tätigkeiten soll der Begriff Textverwaltung hier stehen. Zwar wird in all diesen Fällen auch geschrieben und gelesen, aber das ist nicht das Wesentliche an der Textverwaltung, sondern die in diesen Tätigkeiten erzeugten Ordnungen oder Relationen.

Fragen wir uns nun, zu welchen dieser drei Tätigkeitsfeldern (Lesen, Schreiben, Verwalten) eine digitale Edition eine unmittelbare und eine mittelbare Unterstützung liefern kann. Traditionell dienen digitale Editionen (wie ihre gedruckten Vorfahren) in erster Linie dazu gelesen zu werden. Zwar sind die in ihr enthaltenen Texte und ihre Metadaten natürlich auch Ergebnis einer Textverwaltung. Jedoch bieten sie dem Nutzer nur in seltenen Fällen und da auch nur rudimentär die Möglichkeit selbst Text zu verwalten (cf. z. B. Arbeitsmappen bei Jung 2015). Eine außergewöhnliche Ausnahme ist ein Editionsprojekt zu Pessoa's „Buch der Unruhe“ (cf. Silva / Portela 2015). Hier ist das Lesen, Schreiben und Verwalten gleichermaßen möglich und ermöglicht den Nutzern aus dem vorhandenen Textmaterial und eigenen Kommentaren eine eigene virtuelle Edition kollaborativ zu erstellen. In diesem Sinne ist diese Plattform nicht mehr eine Edition im traditionellen Sinne, sondern selbst eine in sich abgeschlossene Forschungsumgebung – allerdings für eine ganz spezielle Aufgabe über ein abgegrenztes Textkorpus.

All diesen digitalen Editionen ist jedoch gemeinsam, dass, sofern sie eine Textverwaltung unterstützen, diese dann nur für die im System vorhandenen (oder darin erzeugten) Texte ermöglichen. Im Allgemeinen ist der Textwissenschaftler aber nicht mit einem einzelnen Textkorpus befasst, sondern mit mehreren. Eine Textverwaltung kann dann nur ihren Nutzen entfalten,

wenn sie als eigenständiger Service auf mehrere digitale Editionen zugreifen kann.

Nehmen wir als einfaches Beispiel die Zusammenstellung der Literatur zu einem Germanistikseminar, in dem Texte verschiedener Autoren behandelt werden. Von einer komfortablen Textverwaltung würde man jetzt nicht die URL der jeweiligen digitalen Editionen erwarten, sondern man möchte am besten die Texte selbst per Mausklick zur Verfügung gestellt bekommen ohne dabei auf die Web-Seiten der jeweiligen digitalen Editionen gehen zu müssen. Schon dieser einfache Fall zeigt den Nutzen, den eine programmatische Schnittstelle von digitalen Editionen haben könnte: Ein eigenständiger Service zur Aggregation von Semesterapparaten ließe sich mit geringem Aufwand implementieren.

Tatsächlich bieten manche digitale Editionen (z. B. das Deutsche Textarchiv) ihre Texte (sogar in verschiedenen Formaten: TEI, HTML, plain text) zum Download an, so dass man die entsprechenden Links schon als Web-API auffassen könnte. Allerdings beschränkt sich diese Möglichkeit entweder auf den Download einer einzelnen Seite oder des gesamten Textdokuments. Für eine brauchbare Textverwaltung wäre es jedoch wesentlich praktischer, wenn man Texte nicht nach Paginierungsgrenzen sondern bezüglich semantischer Sinneinheiten beziehen könnte. Es fällt nicht schwer, sich entsprechende Szenarien vorzustellen: Für eine Anthologie möchte man etwa Balladen einer bestimmten Epoche zusammenstellen.; für eine Theaterprobe möchte jeder Schauspieler eine Zusammenstellung derjenigen Szenen, in der seine Rolle vorkommt; ein Übersetzungsforscher möchte alle deutschen Übersetzungen des Monolog der ersten Szene im dritten Aufzug von Shakespeares Hamlet. Die Zahl weiterer Szenarien ist unbegrenzt. Als entscheidende Anforderung an eine digitale Edition wäre festzuhalten: die Adressierbarkeit und Auffindbarkeit von Texten in allen üblichen Struktureinheiten (z. B. Kapitel, Absatz, Drama, Akt, Szene, Gedicht, Strophe, Vers, etc.). Da in den meisten digitalen Editionen die Texte im TEI-XML vorliegen, welche die Kodierung solcher Struktureinheiten erlauben, dürfte es prinzipiell nicht schwierig sein, diese auch über eine Web-API adressierbar zu machen. Was die Auffindbarkeit betrifft, wäre es wünschenswert, die Möglichkeitender in der Backend-Datenbank verwendeten Anfragesprachen weitgehend in der Web-API abzubilden. Das ganze Feld der Suchmöglichkeiten ist allerdings so umfangreich, dass es einen eigenen Beitrag rechtfertigen würde und daher hier nicht weiter vertieft werden soll. Allein die Adressierbarkeit aller textspezifischen Struktureinheiten (s. o.) mittels der Web-API von digitalen Editionen wäre eine große Chance zur Entwicklung nützlicher Textverwaltungsdienste. Allerdings sollten neben den vorgegebenen Struktureinheiten auch vom Nutzer frei definierte Textauswahlen von einer digitalen Edition adressierbar sein. Damit soll die verbreitete Praxis, Textausschnitte mit einem Textmarker zu markieren, im digitalen Medium nicht nur die Funktion erhalten, etwas

farblich hervorzuheben, sondern die so ausgezeichneten Textpassagen sollen durch eine generierte Adresse permanent referenzierbar gemacht werden. Damit wäre beispielsweise eine Sammlung von Exzerpten referenzierbar, die ein Benutzer mit einem virtuellen Textmarker erzeugt hat.

Bis hierin wurde die Adressierbarkeit von jeglichen Textausschnitten in den oben angeführten Szenarien ausschließlich für die Erstellung von Textsammlungen verwendet. Das ist aber nur eine einfache Form der Textverwaltung. Denn eine Textsammlung ist zunächst eine in sich unstrukturierte Menge von Texten. Ziel einer Textverwaltung ist es aber meist, in eine Textsammlung eine bestimmte Ordnung zu bringen. Das ist unter anderem der Fall, wenn man die gesammelten Texte nach forschungseigenen Kriterien klassifiziert; z. B. als Linguist nach grammatischen Eigenschaften, als Literaturwissenschaftler nach Motiven, als Übersetzer nach Idiomen, etc.

Textklassifikation wäre eine Relation zwischen Texten und Sammelbegriffen. Darüber hinaus wäre es wichtig, in einer Textverwaltung die Beziehung der Texte untereinander explizit machen zu können. So könnte man beispielsweise explizit erfassen, dass eine bestimmte Textpassage eine Anspielung auf einen anderen Text ist; oder dass die eine Textfassung aus jener Skizze hervorgegangen ist, etc. Soweit würde man Textausschnitte aus digitalen Editionen in Beziehung zueinander setzen. Man würde aber in einer Textverwaltung insbesondere auch die Texte der digitalen Editionen in Beziehung zu selbstverfassten Texten setzen wollen. Auch würde man Texte zu nicht textartigen Gegenständen wie Personen, Orte oder Ereignissen in Beziehung setzen wollen; beispielsweise wenn man in historischen Romanen den Bezug zu historisch belegten Sachverhalten herstellen möchte.

Eine Textverwaltung, die all die skizzierten Funktionalitäten bereitstellen würde, könnte einen Textwissenschaftler bei der Arbeit am Text bzw. der Organisation der eigenen Texte erheblich unterstützen. Sie würde darüber hinaus das kollaborative Arbeiten erleichtern, indem sie eine auf Austausch von Dokumenten basierte Arbeitsweise durch eine Praxis der direkten Vernetzung von Inhalten im Netz ersetzen würde. Sie könnte aber nur funktionieren, wenn die Texte digitaler Editionen in aller Granularität über Web-APIs adressierbar wären.

Abschließend soll erwähnt werden, dass eine ganze Reihe von Anstrengung von verschiedenen Seiten schon unternommen wurden, die durch eine geeignete Zusammenführung ein solides Fundament zur Umsetzung dieser Visionen bilden könnten. Allgemeine technische Grundlage wären die Semantic-Web-Technologien. Darauf aufbauend wären folgende theoretische und praktische Arbeiten hervorzuheben: Von Silvio Peroni (2014) zu „Semantic Publishing“, Fabio Ciottis und Francesca Tomasis (2014) Entwurf zu „Formal ontologies, Linked Data and TEI semantics“, das semantic annotation Tool

Pundit (2013-*) und die Open Annotation Initiative : <http://www.openannotation.org/>.

Bibliographie

Ciotti Fabio, Tomasi Francesca (2014): *Formal ontologies, Linked Data and TEI semantics*. TEI Conference and Members Meeting 2014. Evanston (IL), October 22-24, 2014. <http://tei.northwestern.edu/files/2014/10/Ciotti-Tomasi-22p2xtf.pdf> [letzter Zugriff 09. Januar 2016].

Jung, Joseph (ed.) (2015): *Digitale Briefedition Alfred Escher*. Version: Juli 2015. Zürich: Alfred Escher-Stiftung. <http://www.briefedition.alfred-escher.ch/> [letzter Zugriff 09. Januar 2016].

Peroni, Silvio (2014): *Semantic Web Technologies and Legal Scholarly Publishing*. Switzerland: Springer International Publishing <http://www.springer.com/us/book/9783319047768> [letzter Zugriff 09. Januar 2016].

Pundit (2013-*): *Pundit net7* <http://thepund.it/> [letzter Zugriff 09. Januar 2016].

Sahle, Patrick (2013): *Digitale Editionsformen*. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels: Befunde, Theorie und Methodik (= Schriften des Instituts für Dokumentologie und Editorik 8). Norderstedt: Books on Demand.

Silva, António Rito / Portela, Manuel (2015): "TEI4LdoD: Textual Encoding and Social Editing in Web 2.0 Environments", in: *Journal of the Text Encoding Initiative* 8 <http://jtei.revues.org/1171> [letzter Zugriff 09. Januar 2016].