

SEMANTIC INDEXING OF FIGURE CAPTIONS

Dr. Joachim Brand^a, Frank Dührkohp^b

^aKunstabibliothek, Staatliche Museen zu Berlin – Preußischer Kulturbesitz, Deutschland, j.brand@smb.spk-berlin.de; ^bDigitale Bibliothek, Verbundzentrale des Gemeinsamen Bibliotheksverbundes, Göttingen, Deutschland, Frank.Duehrkohp@gbv.de

KURZDARSTELLUNG: Die Kunstbibliothek, Staatliche Museen zu Berlin und die Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG; Göttingen) planen gemeinsam mit den Technikpartnern IntraData GmbH (Göttingen), Programmfabrik GmbH (Berlin) und qouch (Jena) ein Forschungsprojekt zur Extraktion, automatischen Indexierung und Präsentation von Abbildungen aus digitalisierten Druckmedien in Repositorien. Das Projekt soll die folgenden Ergebnisse liefern: 1) Eine Bilddatenbank (easyDB) mit den semantisch erschlossenen Abbildungen und ein Repository (IntraDataViewer/Goobi) mit den semantisch angereicherten Volltexten der Pilotzeitschrift "Die Kunst für alle". 2) Einen nachnutzbaren Webservice zur automatischen semantischen Indexierung und Extraktion von Informationen zu Abbildungen für die spätere Erschließung weiterer Repositorien. 3) Eine Ressource und einen Webservice für das im Pilotprojekt extrahierte Fachvokabular.

1. PROBLEMSTELLUNG

Digitalisierte Bücher und Zeitschriften enthalten in ihren Abbildungsteilen eine Fülle relevanter Abbildungen für die Kunstgeschichte und andere Wissenschaften. Insbesondere historische Kunstzeitschriften haben für die Kunstwissenschaft einen immensen Quellenwert. Sie zeigen Kunstwerke, die nicht mehr existieren, die in ihrem Zustand verändert wurden, oder die sich in unzugänglichen Privatsammlungen befinden. Leider sind diese wichtigen Bildquellen nur sehr eingeschränkt für die wissenschaftliche Auswertung und Forschung verfügbar, weil die derzeitigen Methoden zur Recherche nach Bildinhalten allesamt unzureichend sind. Recall und precision des klassischen *Text-based image retrieval* in speziellen Bilddatenbanken wird bei der Suche nach Bildern in einem Repository weder von der in der Regel verfügbaren Volltextsuche noch von der Bildähnlichkeitssuche erreicht. In der hinsichtlich Bilderkennungsverfahren führenden medizinischen Informatik findet sich eine vergleichbare Bewertung der Verfahren: "In the domain of medical imaging informatics, a huge amount of image data is being produced. A lot of work

has already been done to improve the image retrieval systems. One is text-based approach and the other is content-based. Each method has its own advantages and disadvantages. Text-based method is widely used and fast, but it requires precise annotation. Content-based approach provides semantic retrieval, but effective and precise techniques still remains elusive." [1] Dieses sogenannte *Content-based image retrieval* (CBIR) basiert auf dem Vergleich von Merkmalen (Helligkeit, Textur, Umriss, Farbe usw.) der Bilder selbst und findet sich in Repositorien derzeit nur im Versuchsstadium. Die Bayerische Staatsbibliothek hat für mehrere Millionen Bilder aus ihrem Druckschriftenrepository eine CBIR-basierte Suchmöglichkeit implementiert (<http://bildsuche.digitale-sammlungen.de/>). [2] Das Verfahren kann überraschende Treffer liefern. Es ist zu einer systematischen Suche jedoch (noch) nicht geeignet. "Though many sophisticated algorithms have been designed to describe color, shape, and texture features, these algorithms cannot adequately model image semantics and have many limitations when dealing with broad content image databases. Extensive experiments on CBIR systems show

that low-level contents often fail to describe the highlevel semantic concepts in user's mind. Therefore, the performance of CBIR is still far from user's expectations.”[3] Die einzige derzeit verfügbare Alternative zu den beiden für systematische Suchen in Repositorien unzureichenden Verfahren ist das intellektuelle Tagging von Bildunterschriften in den METS/MODS-Strukturdaten der Zeitschrifteninhalte, das jedoch sehr aufwändig und damit teuer ist.

2. PROJEKTZIELE

Im Forschungsprojekt *Semantic Indexing of Figure Captions (SIFC)* werden die Texte einer digitalisierten Publikation und die in ihnen enthaltenen Abbildungen anstatt wie derzeit als eine Quelle, die mittels Volltextindexierung erschlossen ist, als zwei eigenständige Quellen betrachtet, die jeweils eigener Erschließungsverfahren mit unterschiedlichen Workflows bedürfen. Im Projekt SIFC soll am Beispiel der Zeitschrift „Die Kunst für Alle“ ein weitgehend automatisiertes Verfahren zur Verbesserung der Suche nach digitalen Abbildungen in Repositorien entwickelt werden. Die von 1885-1942 erschienene Zeitschrift „Die Kunst für Alle“ ist ein geeignetes Pilotobjekt für das Verfahren, weil sie auf rund 25.000 Seiten ca. 30.000 bis 35.000 Abbildungen enthält, die sämtlich mit weitgehend normierten bzw. normierbaren Bildunterschriften versehen sind.



Abb. 1: Die Kunst für Alle



Abb. 2: Abbildungen mit Bildlegenden

Basis des neuen Verfahrens sind die praktisch in allen Arten von Publikationen vorliegenden Bildlegenden. Bei diesen Bildunterschriften handelt es sich um eine spezielle Sorte von Paratexten mit syntaktisch und semantisch hoch spezifischen Funktionen. Sie sind damit eine für computerlinguistische Analysen sehr gut geeignete Textklasse. Die Projektskizze folgt einem Ansatz, der in den USA erfolgreich zur Verbesserung der Bildrecherche in Bilddatenbanken in der Radiologie eingesetzt wird.[4] Bildlegenden bestehen aus einer Reihe von distinkten Elementen, die das Bild in unterschiedlichen Bezugssystemen verankern. Diese ancrage-Funktion nach Roland Barthes ist für alle Arten von Bildunterschriften konstitutiv. Zentrale Bezugssysteme im Bereich Kunst und Kunstwissenschaft sind die kunsthistorischen Konzepte Künstler, Gattungen und Sammlungen (Besitzer), die Fakten Technik, Maße und Entstehungsjahr, ggf. Metadaten zu den abgebildeten Objekten wie Signaturen, Referenzen (bspw. Literaturangaben) und Fotografennamen sowie rechtlich notwendige Daten wie Copyrightvermerke oder Creditlines. Diese sinntragenden Elemente einer Bildlegende sind eine geeignete Basis für eine Datenstruktur, auf der eine Suche nach Bildinhalten in einem Repositorium aufsetzen kann. Aufgrund der vorhandenen Mehrdeutigkeiten in Bildunterschriften werden ergänzend im Analyseprozess auch die

Volltexte der „Kunst für Alle“ linguistisch-semantic analysiert. Wenn *SIFC* die erwarteten Ergebnisse liefern kann, dann werden künftig spezifische Suchen nach Künstlern, Sammlungen, Werktiteln, Gattungen und künstlerischen Techniken möglich, die derzeit für die Bildinhalte von Repositorien nicht möglich sind.

3. ARBEITSSCHRITTE

Im Projekt *Semantic Indexing of Figure Captions* werden die Texte einer digitalisierten Publikation und die in ihnen enthaltenen Abbildungen als zwei eigenständige Quellen betrachtet. Ihre Erschließung erfolgt dementsprechend mit unterschiedlichen Verfahren in zwei Workflows. Der Workflow Bilderschließung erfolgt in 8 Arbeitsschritten:

3.1 OCR

Mittels eines speziell parametrisierten OCR-Laufs werden die Bilder und Bildunterschriften aus „Die Kunst für Alle“ als Block selektiert, aus dem Seitenlayout ausgeschnitten und mit den zugehörigen Positionsangaben und bibliografischen Metadaten „Zeitschriftentitel“, „Bandangabe“, „Jahr“ und „Seitenzahl“ im Format ALTO-XML 3.0 abgelegt.

XML-ALto aus OCR

```

...
<ComposedBlock ID="P0048_C001" HPOS="880" VPOS="360" WIDTH="1307" HEIGHT="823">
  <Illustration ID="P0048_IL01" HPOS="880" VPOS="360" WIDTH="1307" HEIGHT="774"
  IDNEXT="P0048_T005" FILEID="048_IL01.jpg"/>
  <TextBlock ID="P0048_T005" HPOS="880" VPOS="1153" WIDTH="1307" HEIGHT="30"
  IDNEXT="P0048_T006">
    <TextLine ID="P0048_TL44" HPOS="880" VPOS="1153" WIDTH="372" HEIGHT="30">
      <String ID="P0048_ST213" CONTENT="Hubert" HPOS="880" VPOS="1153" WIDTH="89" HEIGHT="30"/>
      <String ID="P0048_ST214" CONTENT="Netzer" HPOS="1051" VPOS="1153" WIDTH="201"
      HEIGHT="30"/>
    </TextLine>
    <TextLine ID="P0048_TL45" HPOS="1478" VPOS="1153" WIDTH="709" HEIGHT="30">
      <String ID="P0048_ST215" CONTENT="Der" HPOS="1478" VPOS="1153" WIDTH="89" HEIGHT="30"/>
      <String ID="P0048_ST216" CONTENT="Normenbrunnen" HPOS="1567" VPOS="1153" WIDTH="360"
      HEIGHT="30"/>
      <String ID="P0048_ST217" CONTENT="in" HPOS="1927" VPOS="1153" WIDTH="57" HEIGHT="30"/>
      <String ID="P0048_ST218" CONTENT="Nünchen" HPOS="1984" VPOS="1153" WIDTH="203"
      HEIGHT="30"/>
    </TextLine>
  </TextBlock>
</ComposedBlock>
...

```

Abb. 3: XML-ALto

3.2 PREPROCESSING

Das folgende Preprocessing der Bildlegenden dient dazu, die Datenqualität zu verbessern, strukturelle Mängel aus der OCR-Erkennung zu beheben und die Daten als fortlaufende Texte aufzubereiten. Während des Preprocessings wird ein internes XML-Zwischenformat generiert, in dem ID und Zeichenpositionen der Strings aus XML-ALto abgelegt sind. Mittels Parsing werden Tags und Attribute sowie Zeilen- und Seitenumbrüche aus XML-ALto extrahiert, so dass die Textinhalte im XML-Zwischenformat als Fließtext von der NLP-Pipeline weiterverarbeitet werden können.

XML-Zwischenformat mit Annotationen

```

...
<feature type="com.qouch.parser.ParserResult:POS">[N|ae|KOM|, schon|ADV, mitgeteilt|V|PP, /;S,
wurde|VAFIN, kürzlich|ADV, im|APP, Essen|NM, der|ART, von|APP, dem|ART, Münchener|ADJA, Bildhauer|NM,
Ulfer|NE, Janssen|NE, geschaffene|ADJA, Hundertjahr-Brunnen|NM, enthält|V|FIN, /;S, wir|PPER, geben|
V|FTN, von|APP, dem|ART, wohlgelesenen|ADJA, Werk|NM, untenstehend|APP, eine|ART, Abbildung|NM, /;S,
sich|PPER, zeigt|V|FTN, /;S, wie|P|adv, glücklich|ADJ, dem|ART, Künstler|NM, die|ART, Raumverhältnisse|
NM, des|ART, Platzes|NM, benutzt|V|PP, hat|V|AFIN, und|KON, wie|P|adv, gut|ADJ, ihm|PPER, die|ART, Ver-
bindung|NM, von|APP, Plastik|NM, und|KON, Architektur|NM, gelungen|V|PP, ist|V|AFIN, /;S.]/</feature>
<feature type="com.qouch.parser.ParserResult:Dependency"/>
<feature type="com.qouch.parser.ParserResult:sentenceid">18615d31-3282-4567-b3a0-20e455bb470c/
</feature>
</annotation>
</Annotation type="com.qouch.parser.ParserResult">
<feature type="com.qouch.Annotation:begin">2307</feature>
<feature type="com.qouch.Annotation:end">2485</feature>
<feature type="com.qouch.parser.ParserResult:TextBlock">Die Hauptfigur, ein ruhender Arbeiter, er-
gibt in ihrer Monumentalität einen glücklichen Gegensatz zu den in den Öffnungen der Brunnenwand ange-
brachten, lustig spielenden Putten. </feature>
<feature type="com.qouch.parser.ParserResult:Tree">
</feature>
</ROOT>
</NP-SB>
</ART>
Die
</ART>
...

```

Abb. 4: XML-ALto mit Annotationen

3.3 LINGUISTISCH-SEMANTISCHE ANALYSE

Die Plattform qouch® Analytics zur linguistisch-semantic Analyse unstrukturierter Daten besteht aus verschiedenen NLP-Komponenten und Webservices, die in Pipelines organisiert sind. Grundlage für die Named Entity Recognition sind die Entitätenkategorien der GND (Personen, Körperschaften, Geografika, Werke und Veranstaltungen). Ergebnis der linguistisch-semantic Analyse sind extrahierte Named Entities sowie Terme (Sachbegriffe).

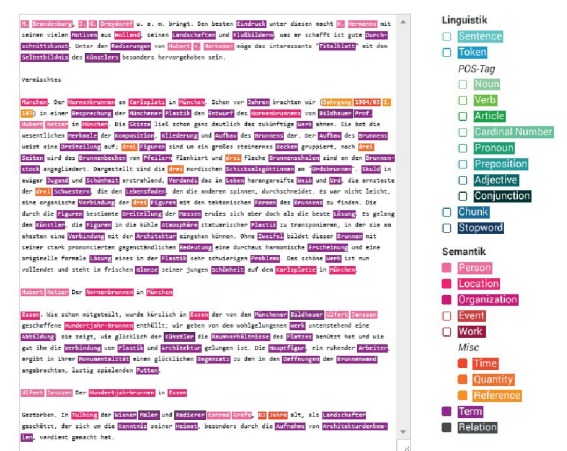


Abb. 5: Linguistisch-semantic Analyse

3.4 CONCEPT MAPPING

Die aus Volltexten und Bildunterschriften extrahierten Entitäten und Sachbegriffe werden mittels semantischer Technologien in eine Fachterminologie überführt und über Mapping-Algorithmen mit Normdatenressourcen (GND; dbpedia) und den Konzepten der Vokabularien des Getty Research Institutes (AAT, TGN, ULAN) vernetzt. Die Getty-Vokabularien sind insbesondere für die Kunstwissenschaft von großer Bedeutung.

Tag-ID	Tag-Label	Concept	Kategorie
TAG_27	Hubert Netzer	http://d-nb.info/gnd/110944043	Person
TAG_67	Ulrich Janssen	http://d-nb.info/gnd/118711814	Person
TAG_83	Konrad Grefe	http://d-nb.info/gnd/11682269X	Person
TAG_15	München	http://d-nb.info/gnd/4127799-4	Location
TAG_16	Nornenbrunnen	http://d-nb.info/gnd/4127799-4	Location
TAG_17	Karlsplatz	http://d-nb.info/gnd/4021938-4	Location
TAG_66	Essen	http://d-nb.info/gnd/4003557-2	Location
TAG_68	Hundertjahn-Brunnen	http://d-nb.info/gnd/4128085-2	Location
TAG_79	Tulding	http://d-nb.info/gnd/4128085-2	Location
TAG_26	drei	http://d-nb.info/gnd/4128085-2	Quantity
TAG_19	Jahrgang 1994/96 S. 187	http://digi.ub.uni-heidelberg.de/diglit/vfaissw_1994/0187	Reference
TAG_14	Künstler	http://d-nb.info/gnd/4002032-5	Term
TAG_18	Jahr	http://d-nb.info/gnd/4182925-9	Term
TAG_24	Plastik	http://d-nb.info/gnd/4004277-8	Term
TAG_24	Entwurf	http://d-nb.info/gnd/4121238-3	Term
TAG_26	Bildhauer	http://d-nb.info/gnd/4129919-5	Term
TAG_26	Professor	http://d-nb.info/gnd/4002503-7	Term
TAG_28	Skizze	http://d-nb.info/gnd/4181468-7	Term
TAG_28	Werk	http://d-nb.info/gnd/4117463-9	Term
TAG_31	Komposition	http://d-nb.info/gnd/4002892-1	Term
TAG_34	Brunnen	http://d-nb.info/gnd/4008491-7	Term
TAG_37	Figur	http://d-nb.info/gnd/4122928-1	Term
TAG_43	Pfeiler	http://d-nb.info/gnd/4174814-2	Term
TAG_44	Schicksalsgöttin	http://d-nb.info/gnd/4124143-5	Term
TAG_45	Urdsbrunnen	http://d-nb.info/gnd/4124143-5	Term
TAG_46	Skuld	http://d-nb.info/gnd/4124143-5	Term
TAG_47	Jugend	http://d-nb.info/gnd/4002899-4	Term
TAG_48	Schönheit	http://d-nb.info/gnd/4003125-9	Term
TAG_49	Vernandi	http://d-nb.info/gnd/4003125-9	Term
TAG_50	Leben	http://d-nb.info/gnd/4004893-3	Term
TAG_52	Urd	http://d-nb.info/gnd/4004893-3	Term
TAG_54	Lebensfaden	http://d-nb.info/gnd/4124142-2	Term
TAG_60	Architektur	http://d-nb.info/gnd/4002893-3	Term
TAG_69	Abbildung	http://d-nb.info/gnd/4125432-1	Term
TAG_71	Platz	http://d-nb.info/gnd/4004238-8	Term
TAG_73	Arbeiter	http://d-nb.info/gnd/4125093-8	Term
TAG_74	Monumentalität	http://d-nb.info/gnd/4029296-9	Term
TAG_76	Öffnung	http://d-nb.info/gnd/4179393-8	Term
TAG_78	Putz	http://d-nb.info/gnd/4148824-9	Term
TAG_81	Heiler	http://d-nb.info/gnd/4007315-7	Term
TAG_82	Redierer	http://d-nb.info/gnd/4176793-7	Term
TAG_87	Heinet	http://d-nb.info/gnd/4004105-1	Term
TAG_89	Architekturdenkmal	http://d-nb.info/gnd/4003403-3	Term

Abb. 6: Concept Mapping

3.5 POSTPROCESSING

Die Ergebnisse von Analyse und Concept Mapping werden im XML-Zwischenformat ausgegeben. Das Postprocessing dient dazu, diese Ergebnisse im XML-AltO mittels der hinterlegten IDs und Zeichenpositionen der Strings zusammenzuführen. Das Resultat des Postprocessing ist ein annotierter XML-AltO-File.

XML-AltO mit Annotationen und Konzeptlinks

```

<NamedEntityTag ID="TAG_01A" LABEL="Martin Brandenburg"
  URI="http://de.dbpedia.org/resource/Martin_Brandenburg"/>
<NamedEntityTag ID="TAG_04A" LABEL="Heinrich Hermanns"
  URI="http://de.dbpedia.org/resource/Heinrich_Hermanns"/>
<NamedEntityTag ID="TAG_05A" LABEL="Netsi"
  URI="http://de.dbpedia.org/resource/Netsi_(bildende_Kunst)"/>
<NamedEntityTag ID="TAG_06A" LABEL="Holland"
  URI="http://de.dbpedia.org/resource/Niederlande"/>
<NamedEntityTag ID="TAG_07A" LABEL="Landschaft"
  URI="http://de.dbpedia.org/resource/Landschaft"/>
<NamedEntityTag ID="TAG_10A" LABEL="Radierung"
  URI="http://de.dbpedia.org/resource/Radierung"/>
<NamedEntityTag ID="TAG_11A" LABEL="Hubert von Herkener"
  URI="http://de.dbpedia.org/resource/Hubert_von_Herkener"/>
<NamedEntityTag ID="TAG_12A" LABEL="Titelblatt"
  URI="http://de.dbpedia.org/resource/Titelblatt"/>
<NamedEntityTag ID="TAG_13A" LABEL="Künstler"
  URI="http://de.dbpedia.org/resource/Selbstbildnis"/>
<NamedEntityTag ID="TAG_14A" LABEL="Künstler"
  URI="http://de.dbpedia.org/resource/Künstler"/>
<NamedEntityTag ID="TAG_15A" LABEL="München"
  URI="http://de.dbpedia.org/resource/München"/>
<NamedEntityTag ID="TAG_16A" LABEL="Nornenbrunnen"
  URI="http://de.dbpedia.org/resource/Nornenbrunnen_(München)"/>
<NamedEntityTag ID="TAG_17A" LABEL="Karlsplatz" URI="http://de.dbpedia.org/resource/Stachus"/>
<NamedEntityTag ID="TAG_18A" LABEL="Jahr" URI="http://de.dbpedia.org/resource/Jahr"/>
<NamedEntityTag ID="TAG_21A" LABEL="Besprechung" URI="http://de.dbpedia.org/resource/Besprechung"/>
<NamedEntityTag ID="TAG_23A" LABEL="Plastik" URI="http://de.dbpedia.org/resource/Plastik_(Kunst)"/>

```

Abb. 7: XML-AltO Endformat

3.6 EVALUATION SEMANTISCHE INDEXIERUNG

Die Evaluation der Ergebnisse von Analyse und Mapping erfolgt in einem Webservice. In diesem sind die aus den Bildlegenden extrahierten Konzepte, die zugehörigen

Texte der Bildlegenden und die vorhandenen Konzepte mit den zugehörigen Termen aus der GND gelistet.

3.7 AUFBAU DER BILDDATENBANK

Die von gouch generierten Metadaten für die Einzelabbildungen liegen als RDF/XML-Repräsentation vor und werden an die Anforderungen der Datenbank easydb angepasst. Die zugehörigen Bilder werden anschließend mit den Metadaten in die Bilddatenbank easyDB eingelesen und indexiert.

3.8 PRÄSENTATIONSKONZEPTE

Die Präsentation und Recherche erfolgt in der generierten Bilddatenbank im Suchinterface von easyDB (Programmfabrik GmbH). Die von gouch generierten Metadaten werden in Repositorien auf Basis des IntranداViewers bzw. Goobis integriert (Intranدا GmbH). Alternative Webservices und Visualisierungen (Viewer, Timeline, GeoMap, ConceptMap) werden bei Bedarf durch gouch, Intranدا und Programmfabrik entwickelt.

The screenshot shows a web interface for the Nornenbrunnen in Munich. At the top is a photograph of the fountain. Below it is a metadata panel with the following fields:

- Titel:** Der Nornenbrunnen in München
- Subjekt:** Nornenbrunnen [W](#)
- Beziehungen zu Personen:** Bildhauer: Netzer, Hubert (1865 - 1939) [GND](#) [W](#)
- Typ:** Bauwerk (gib)
- Zeit:** erstellt: 1907
- Geografischer Bezug:** Ort: Karlsplatz, München [GND](#) [W](#)
- Land:** Deutschland [GND](#) [W](#)
- Koordinaten:** E 011°34'18" / N 048°08'33" E 011.571719 / N 048.142689
- Oberbegriffe:** Beispiel für: Brunnen [GND](#) [W](#)
- Themen:** 13.2 Plastik; 31.3a Architektur
- Verwandte Abbildungen:**
- Sachbegriffe:** Dreiteilung, Schicksalsgöttin, Urdsbrunnen, Skuld, Vernandi, Urd, Lebensfaden
- Thema in:** Kunst für alle - Band 23 - Vermischtes - München

Abb. 8: Alternativer Viewer mit semantischen Links

4. DANKSAGUNG

Wir danken Steffen Hankiewicz und Robert Sehr (Intranda GmbH, Göttingen), Sebastian Klarmann und Dominika Pienkos (Programmfabrik GmbH, Berlin), Konstantin und Kathrin Hotzel (qouch, Jena) und Andreas Bienert für anregende Diskussionen zum Projekt. Kathrin und Konstantin Hotzel gilt ein besonderer Dank für die Konzeption und Beschreibung wesentlicher Teile des geplanten Workflows.

5. LITERATURHINWEIS

1. Kyung Hoon Hwang et al.: Medical Image Retrieval : Past and Present, in: *Healthcare Informatics Research*. 2012 March;18(1):3-9., S. 8
<http://dx.doi.org/10.4258/hir.2012.18.1.3>
2. Bildersuche in digitalen Massendaten, in: *ABI Technik* 2013; 33(2): 104–108. DOI 10.1515/abitech-2013-0015.
3. Ying Liu et al.: A survey of content-based image retrieval with high-level semantics, in: *Pattern Recognition* 40 (2007) 262 – 282, S.263.
4. Kahn, Charles E., Rubin, Daniel L.: Automated Semantic Indexing of Figure Captions to Improve Radiology Image Retrieval, in: *Journal of the American Medical Informatics Association*. Volume 16 Number 3 May / June 2009, S. 380 –386. DOI 10.1197/jamia.M2945.