

Cheval: Das Churer Evaluationskonzept für Information, Systeme und Nutzer - eine Toolbox im Entstehen.

Josef Herget, Thomas Weinhold, Sonja Hierl, Lydia Bauer, Joachim Pfister
Schweizerisches Institut für Informationswissenschaft (SII)
Hochschule für Technik und Wirtschaft (HTW) Chur, Schweiz
josef.herget@fh-htwchur.ch
thomas.weinhold@fh-htwchur.ch
sonja.hierl@fh-htwchur.ch
lydia.bauer@fh-htwchur.ch
joachim.pfister@fh-htwchur.ch

Abstract: Die Bewertung von Kontexten der Informationsarbeit bleibt eine unverzichtbare Aufgabe, um effektive und effiziente Arbeitsumgebungen zu gestalten. Evaluationskonzepte erstrecken sich auf die Information selbst sowie die sie verarbeitenden Systeme, umfassen aber auch die Verwendung und Verwertung durch die Nutzer. Gerade die Evaluation von Information Retrieval Systemen (IRS) kann wesentlich dazu beitragen, die Kosten für die Informationsbeschaffung zu reduzieren, da hierdurch die Identifikation von Systemen ermöglicht wird, die möglichst optimal an die jeweilige Anwendungsumgebung und die jeweiligen Bedürfnisse angepasst sind. Während die Evaluation klassischer Information Retrieval Systeme und Suchmaschinen auf standardisierten Prozessen und Kennzahlen beruht, haben sich für die Evaluation von IRS, die neue Konzepte, wie z.B. Visualisierungskomponenten oder Ansätze des Semantic Web berücksichtigen, bisher keine solchen Standards etabliert. Diese Lücke möchte das Schweizerische Institut für Informationswissenschaft (SII) mit der Entwicklung des „Chur Evaluation Laboratory“ (CHEVAL) schliessen. Das Ziel von CHEVAL ist es, eine integrierte, mehrstufige und mehrdimensionale, webbasierte Evaluationsumgebung zu entwickeln, die unterschiedliche Evaluationsmethoden aus den Bereichen der IR-Effektivitätsmessung, dem Usability-Testing, der Nutzerbefragung und der Wirtschaftlichkeitsanalyse unterstützt. Der vorliegende Beitrag gibt zunächst einen kurzen Einblick in die Problematik der Beurteilung des Wertes von Informationen. Anschliessend wird auf bestehende Initiativen zur Evaluation von IRS eingegangen. Im Hinblick auf die bestehenden Defizite, die diese Initiativen in Bezug auf die Evaluation von IRS, die neue Konzepte beinhalten, aufweisen, werden die Vision und die Ziele von CHEVAL vorgestellt. Anschliessend wird die für die Testumgebung konzipierte Systemarchitektur vorgestellt und durch zwei Anwendungsfälle illustriert. Den Abschluss bilden ein Ausblick auf potentielle Erfolgsfaktoren hinsichtlich der Entwicklung der Testumgebung, sowie ein kurzes Fazit.

Schlagwörter: Evaluation, Information Retrieval Systeme, Usability, Wirtschaftlichkeitsanalyse

1 Information in der Wirtschaft und die Wirtschaftlichkeit von Information

Wirtschaftliches Handeln bezweckt eine optimale Ressourcenallokation mit bestmöglicher Resultaterzielung und vollzieht sich in den Schritten Zielsetzung, Planung, Entscheidung, Durchführung und Kontrolle. Unter Berücksichtigung der Umweltbedingungen von Entscheidungssituationen entsteht eine enge Verbindung zwischen Entscheidung und Information (Schoop 2004). Der Wert einer Information steht also in einem engen Bezug zu ihrer Verwendung und nicht zu ihrer Produktion. Informationen haben daher keinen absoluten Wert, sondern ihr Wert ist immer in Abhängigkeit von ihrem Gebrauch zu sehen (Krcmar 2005). Die Quantifizierung des Wertes von Information bleibt daher schwierig. Hingegen ist es durchaus möglich, die Kosten, die in Zusammenhang mit der Informationsbeschaffung, Informationsspeicherung, Informationsverarbeitung und Informationsweitergabe stehen, zu eruieren. Gerade in Bezug auf die Informationsbeschaffung sind in den letzten Jahren vor dem Hintergrund des Erfolges des Internets grosse Veränderungen festzustellen. Die Nutzer stehen heute einem nahezu unbegrenzten Angebot an verfügbaren Informationen gegenüber. Bedingt durch die immense Masse an Dokumenten wird es allerdings für die Nutzer auch immer schwieriger, die für sie relevanten Informationen aus der Gesamtheit herauszufiltern. Deshalb ist es nicht verwunderlich, dass Suchmaschinen bei Untersuchungen regelmässig zu den am häufigsten genutzten Websites und Softwareapplikationen zählen.

Aus diesem Grund kommt Suchmaschinen und anderen Information Retrieval Systemen (IRS) in Bezug auf einen wirtschaftlichen Umgang mit Informationen eine besondere Bedeutung zu. Die eingesetzten Systeme sollten möglichst gut an die Bedürfnisse ihrer Nutzer angepasst sein. Um dies sicherzustellen, bedarf es der Evaluation der eingesetzten bzw. der potentiell einsetzbaren Systeme. Durch eine Evaluation können die für die jeweilige Systemumgebung und den konkreten Informati-

onsbedarf der jeweiligen Mitarbeiter am besten geeigneten Systeme ermittelt werden, wodurch sich die Kosten der Informationsbeschaffung positiv beeinflussen lassen.

2 Evaluation von Information Retrieval Systemen

Die Evaluation von IRS blickt, ausgehend von den Cranfield-Experimenten in den 60er Jahren, bereits auf eine lange Tradition zurück. Die Cranfield-Experimente verdeutlichten die Bedeutung der Indexierung und der darauf aufbauenden Suche von Texten. Zudem haben diese Tests dazu beigetragen, dass Bewusstsein für die Notwendigkeit der Entwicklung von Testkollektionen für vergleichende Evaluationen zu erweitern (Harman 1993).

Vor diesem Hintergrund wurde 1992 die erste „Text Retrieval Conference“ (TREC) durchgeführt. TREC wurde mit der Absicht geschaffen, eine offene Evaluationsveranstaltung mit einheitlichen Aufgaben und standardisierten Evaluationsszenarien anzubieten, die zu einer Verbesserung und Beschleunigung der Forschung auf dem Gebiet der Evaluation von IRS beitragen sollte. Ausgehend von diesem Grundgedanken haben sich in den letzten Jahren eine Reihe weiterer Evaluierungsinitiativen, wie z.B. das „Cross Language Evaluation Forum“ (CLEF) oder die „Initiative for the Evaluation of XML Retrieval“ (INEX) etabliert.

Die genannten Initiativen und Projekte haben in hohem Masse zur Standardisierung der Evaluation von IRS beigetragen und sind im Bereich der Informationswissenschaft weithin anerkannt. Bedingt durch die Tatsache, dass es diese Initiativen ermöglichen, die selben Aufgaben mit den selben Datensätzen auf unterschiedlichen Systemen durchzuführen, eignen sie sich sowohl für Benchmarking-Zwecke, als auch für wiederholende Studien, die dazu genutzt werden können, ein bestimmtes System kontinuierlich weiter zu verbessern. Zusätzlich stellen diese Initiativen ein Forum bereit, in dem eine grosse Gemeinschaft von Fachleuten neue Retrievaltechnologien diskutieren kann, um den Wissens- und Technologietransfer zu fördern (Vorhees und Harman 2005).

3 Neue Suchkonzepte erfordern neue Evaluationsmethoden und Testumgebungen

Die vorgestellten Evaluierungsinitiativen konzentrieren sich hauptsächlich auf die Messung von „Recall“, „Precision“ und ähnlichen Werten, daher sind diese Initiativen lediglich für die Evaluation von text-basierten Systemen, bzw. mit Hinblick auf INEX von XML-basierten Systemen geeignet. Problematisch dabei ist die Tatsache, dass in den letzten Jahren zu beobachten ist, dass bei einer steigenden Anzahl von Suchmaschinen und anderen IRS neue Konzepte wie Visualisierungskomponenten oder Ansätze des Semantic Web zum Einsatz gelangen. Ein Beispiel hierfür ist die Visualisierung von Suchbegriffen oder den Suchergebnissen mit Hilfe geeigneter grafischer Elemente, wie sie unter anderem bei Quintura, Grokker und einer Reihe weiterer Systeme verwendet wird (Bekavac et al. 2007). Hierbei stellt sich die Frage, ob diese neuen Ansätze geeignet sind, den Nutzern den Zugang zu für sie relevanten Informationen zu erleichtern.

Während die Evaluation klassischer IRS und Suchmaschinen auf standardisierten Prozessen und Kennzahlen beruht, haben sich für die Evaluation von IRS, die neue Komponenten berücksichtigen, bisher keine Standards dieser Art etabliert. Dabei besteht eine starke Abhängigkeit zwischen den semantischen und visuellen Komponenten solcher Systeme und der Art und Weise wie Benutzer mit diesen interagieren, wodurch letztendlich auch die Suchresultate beeinflusst werden. Diese Systeme erfordern daher zur Evaluation unterschiedlicher Ansätze und Methoden (Bauer et al. 2007). Momentan werden entweder standardisierte IR-Effektivitätsmasse oder Masse aus dem Bereich des Usability Testing zur Evaluation von IRS mit Visualisierungskomponenten (Koshman 2005; Reiterer et al. 2005; Zwol und Oostendorp 2004; Reiterer 2004; Sebrechts et al. 1999; Veeresamy und Belkin 1996) verwendet. Jedoch wird in den wenigsten Fällen die Auswahl der gewählten Methoden begründet bzw. eine Kombination von Methoden aus beiden Bereichen eingesetzt. Aus diesem Grund argumentiert Vaughan (2004), dass die Entwicklung zuverlässiger Evaluationsmethoden und –ansätze momentan nicht mit der rasanten Entwicklung von IRS mit Visualisierungskomponenten Schritt halten kann. Ein wesentlicher Aspekt, der bei der Evaluation von IRS die neue, innovative Konzepte beinhalten, berücksichtigt werden muss, sind die Nutzungsgewohnheiten der Anwender, die sich über den jahrelangen Einsatz von klassischen Suchmaschinen wie Google, Yahoo oder MSN etabliert haben (Hierl 2007).

Daher sind bestehende Evaluationskonzepte und –umgebungen dahingehend weiter zu entwickeln, die Untersuchung von IRS in praxisnahen Szenarios, sowie eine Kombination von Methoden aus dem

Bereich der Retrievaleffektivitätsmessung und dem Bereich des Usability Testing zu ermöglichen (Lewandowski und Höchstötter 2007).

4 Beschreibung von CHEVAL

4.1 Vision und Ziele

Die zuvor beschriebenen Probleme bestehender Evaluierungsinitiativen, das Fehlen von integrierten, system-gestützten Ansätzen zur Evaluation von IRS und deren mangelnde Eignung für kleinere, lokale Evaluationen durch Organisationen zur Unterstützung von Auswahlentscheidungen, gaben den Ausschlag für die Konzeption und Entwicklung von CHEVAL.

Die Vision von CHEVAL ist es, eine integrierte, mehrstufige und mehrdimensionale, webbasierte Evaluationsumgebung zu entwickeln, die unterschiedliche Evaluationsmethoden aus den Bereichen der IR-Effektivitätsmessung, dem Usability-Testing und der Wirtschaftlichkeitsanalyse unterstützt. Im Rahmen von CHEVAL kann eine Evaluation mehrere Dimensionen aufweisen, sowohl in Bezug auf den Untersuchungsgegenstand, die Art der Evaluation (Kurzzeit- oder Langzeitstudie, vergleichende oder Nicht-vergleichende Studie, Feldstudie oder Laboruntersuchung), als auch in Bezug auf die zur Evaluation eingesetzten Methoden. Ziel ist es, eine generelle Infrastruktur zur Durchführung von Evaluationen jeglicher Art, wie beispielsweise der Evaluation von Informationssystemen, Usability-Tests oder auch Kunden- und Mitarbeiterbefragungen bereitzustellen. Durch den mehrdimensionalen und multimethodischen Ansatz ist CHEVAL auch zur Evaluation von IRS, die Multimedia- oder Visualisierungskomponenten bzw. Semantische Konzepte oder Ansätze des Web 2.0 beinhalten, geeignet.

Einer der Hauptbeweggründe für die Entwicklung von CHEVAL ist der repetitive Charakter vieler Evaluationen, also der Umstand, dass sich eine Evaluation in der Regel aus mehreren Schritten zusammensetzt, die iterativ für mehrere Testläufe durchgeführt werden müssen. Das System soll dazu beitragen, den Planungsprozess für eine Evaluation und deren Durchführung zu automatisieren und damit zu einer effizienteren Abwicklung von Evaluationen beitragen, wodurch die Kosten für die Durchführung reduziert werden können. Dabei soll vom System die Auswahl geeigneter Evaluationsmethoden und Testkollektionen, die Speicherung der Untersuchungsergebnisse, das Erstellen von Untersuchungsberichten, sowie die Koordination des gesamten Evaluationsprozesses unterstützt werden. Auf diese Weise wird auch Anwendern, die nicht über das notwendige Hintergrundwissen verfügen, um eigene Evaluationskonzepte entwickeln zu können, die Möglichkeit geboten, Evaluationen auf Basis eines standardisierten und wissenschaftlich fundierten Konzepts durchzuführen.

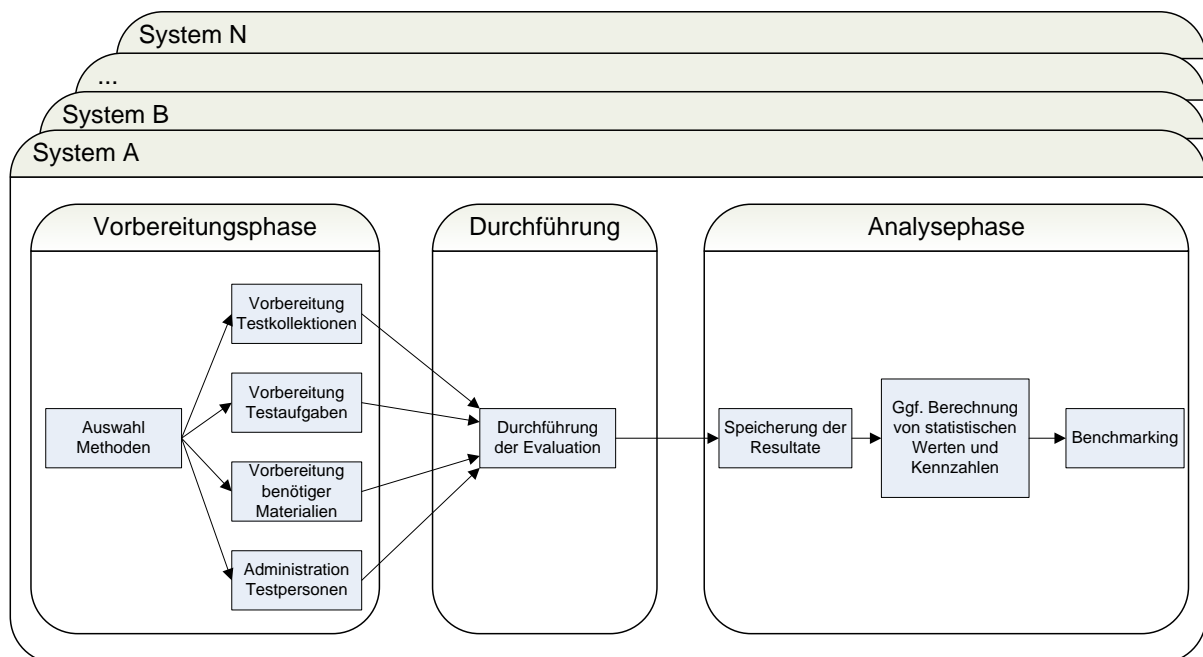


Abbildung 1: Phasen einer Evaluation

Die Testumgebung soll, wie in Abbildung 1 dargestellt, den gesamten Prozess von der Planung einer Evaluation bis zur abschliessen Auswertung und Analyse der Evaluationsergebnisse unterstützen. Mit

Hilfe eines Entscheidungsbaumes soll der Durchführende durch die Vorbereitungsphase der Evaluation geleitet werden. Beispielsweise soll der Entscheidungsbaum einen Anwender bei der Entscheidung unterstützen, ob für das zu evaluierende System eine Kurzzeit- oder Langzeitstudie besser geeignet ist, ob eine vergleichende Analyse durchgeführt werden sollte oder auch welche Methoden für die Evaluation des betreffenden Systems am besten geeignet sind. Alle Untersuchungsergebnisse werden in CHEVAL gespeichert, um weiterführende Analysen zu ermöglichen. Zudem soll es ein Berichtsmodul ermöglichen, unterschiedliche Analysewerkzeuge und –methoden einzusetzen, um die Ergebnisse der Tests aufzubereiten. Bedingt durch die Tatsache, dass alle Evaluationsergebnisse in CHEVAL dokumentiert werden, steht nach einer gewissen Vorlaufzeit eine breite Datenbasis für ein Benchmarking unterschiedlicher Arten von IRS zur Verfügung.

Werden für die Planung oder die Durchführung einer Evaluation zusätzliche, spezifische Informationen benötigt, so können die Nutzer auf die CHEVAL-Wissensbasis (www.cheval-lab.ch) zurückgreifen. Dieser Wissenspool wird in Form eines Wikis bereitgestellt, welches neben Grundlageninformationen zur Evaluation, die unterschiedlichen Evaluationsmasse, -prozesse und –methoden in systematischer Weise vorstellt, die sich beispielsweise zur Evaluation von Suchmaschinen, für Wirtschaftlichkeitsanalysen oder auch für Mitarbeiterbefragungen eignen. Des Weiteren beinhaltet das „Evaluations-Wiki“ Linklisten zur Vertiefung einzelner Themengebiete und bietet somit eine Übersicht über die aktuelle Forschung auf den verschiedenen Evaluationsgebieten. Das Wiki wird sukzessive erweitert und um die während der Evaluationen gewonnen Erkenntnisse durch die CHEVAL Community angereichert. Ausserdem werden Materialien wie beispielsweise Testkollektionen, Heuristiken und Usabilityfragebögen bereitgestellt. Das Wiki bietet somit eine gute Grundlage zur Einarbeitung in das Gebiet der Evaluation, ebenso wie für die Durchführung konkreter Tests.

Ebenso wird in Anlehnung an TREC und andere Evaluierungsinitiativen ein Forum zur Diskussion offener Fragen und interessanter neuer Ansätze aus dem Bereich der Evaluation eingerichtet. Über das Forum soll der Austausch zwischen der Community, den Testpersonen, dem CHEVAL-Entwicklungsteam und weiteren interessierten Personen gefördert werden.

4.2 Systemarchitektur

Mit CHEVAL wird eine Toolbox entwickelt, die Evaluationen in Bezug auf Informationen, Systeme und Nutzer ermöglicht und welche den gesamten Evaluationsprozess elektronisch unterstützt. Die Testumgebung wird als Webapplikation mit einem zentralen Server, auf den die Clients über eine Internetanbindung bzw. das Intranet einer Organisation zugreifen können, implementiert werden. Die Software wird dabei modular aufgebaut sein (vgl. Abbildung 2).

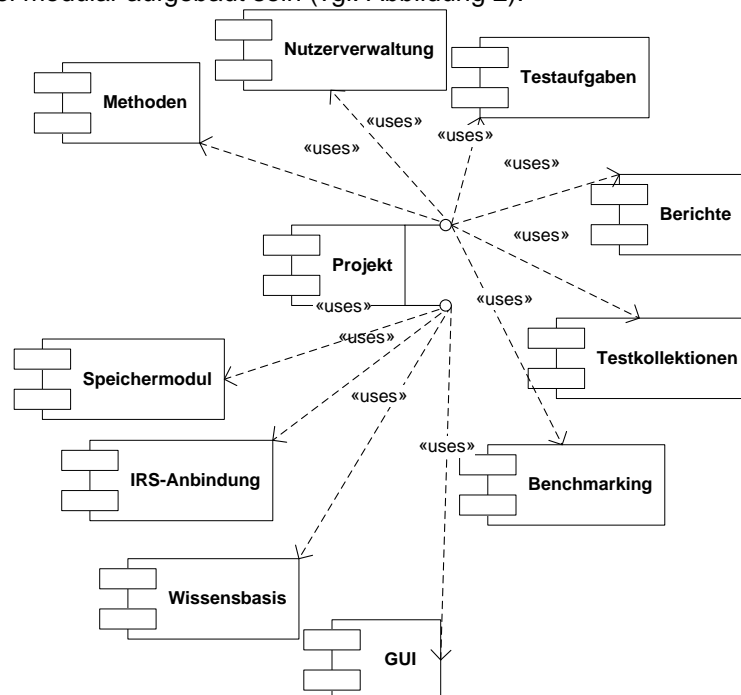


Abbildung 2: Systemarchitektur von CHEVAL

Das Modul „Nutzerverwaltung“ beinhaltet alle Funktionalitäten, die mit der Authentifizierung und der Autorisierung von Benutzern in Zusammenhang stehen. Das System soll ein rollenbasiertes Berechtigungskonzept unterstützen. Dabei kann ein Benutzer folgende Rollen innehaben:

- Systemadministrator: Dieser kümmert sich um die gesamte Testumgebung und stellt deren fehlerfreien und reibungslosen Einsatz sicher. Er ist ausserdem für die Autorisierung weiterer Benutzer und die Modifikation der Testkollektionen verantwortlich.
- Testadministrator: Ein Testadministrator ist für die Einrichtung von Testprojekten zuständig. Er bestimmt, welche Testmethoden und welche Testkollektionen zum Einsatz gelangen sollen. Ausserdem entscheidet er darüber, wer als Testperson zum Einsatz kommt.
- Testperson: Die Testpersonen sind diejenigen, die die geplanten Tests letztendlich durchführen. Hier kann zwischen Experten und Probanden unterschieden werden.

Im Modul „Methoden“ werden die unterschiedlichen Evaluationsmethoden aus den Bereichen Usability-Testing, IR-Effektivitätsmessung und Wirtschaftlichkeitsanalyse, sowie die für diese Methoden benötigten Ressourcen, wie z.B. Fragebögen und ähnliches, verwaltet.

Die Details einer Evaluation werden über das Modul „Projekt“ verwaltet. Dieses kann der Testadministrator nutzen, um die Parameter einer Evaluation zu bestimmen, also beispielsweise welche Testkollektionen und welche Evaluationsmethoden verwendet werden sollen. Im Modul „Testaufgaben“ wird schliesslich bestimmt, welche Aufgaben von den Testpersonen im Rahmen der Evaluation zu bearbeiten sind. Zur Sicherung dieser Angaben, sowie der bei der späteren Durchführung anfallenden Daten, wird das „Speichermodul“ herangezogen. Dort werden alle im Rahmen eines Evaluationsprojektes anfallenden Informationen in einer Datenbank abgelegt.

Nachdem ein Testlauf bzw. eine Evaluation durchgeführt wurde, stellt das Modul „Berichte“ Funktionalitäten für die Erstellung von Statistiken und Berichten zur Verfügung. Das Modul „Benchmarking“ ermöglicht es, die Ergebnisse von einzelnen Testläufen oder auch kompletten Evaluationen miteinander zu vergleichen. Das Modul „Wissensbasis“ stellt einerseits umfassende Informationen zu unterschiedlichen Evaluationsarten und -methoden zur Verfügung und bietet den Anwendern ausserdem die Möglichkeit, weitere Einträge für die Wissensbasis zu erstellen bzw. bereits vorhandene Artikel zu bearbeiten. Dadurch soll dem Testadministrator eine weitere Entscheidungshilfe für die Planung von Evaluationsprojekten gegeben werden. Vervollständigt wird das System durch die GUI-Komponente, die die Schnittstelle zwischen dem System und den Anwendern bildet.

CHEVAL fungiert somit einerseits als Toolbox, die konkrete Systeme und eine Evaluationsumgebung für die Durchführung von Tests zur Verfügung stellt, bei der sich Nutzer die jeweiligen Komponenten entsprechend ihren Bedürfnissen zusammenstellen können. Andererseits bietet CHEVAL Know-How und eine Plattform für den Austausch im Bereich der Evaluationsforschung und -praktiken in der Informationswissenschaft.

4.3 Anwendungsfälle

Um die Anwendungsmöglichkeiten von CHEVAL näher zu beschreiben, werden nachfolgend zwei potentielle Anwendungsfälle dargestellt.

4.3.1 Vergleichende Evaluation von Suchmaschinen mit einer Visualisierungskomponente

Der erste Anwendungsfall beschäftigt sich mit der Evaluation von Suchmaschinen mit Visualisierungskomponenten, wie z.B. Grokker, Quintura oder auch Ujiko. Für die Evaluation solcher Systeme sind Masse wie „Recall“ und „Precision“ nicht geeignet. Vielmehr müssen die Interaktionen der Nutzer mit diesen Systemen bei deren Evaluation mit berücksichtigt werden. Dies wird durch CHEVAL ermöglicht, indem für die Untersuchung der Systeme sowohl Methoden aus dem Gebiet des Usability-Testing als auch Methoden der Retrieval-Effektivitätsmessung zur Verfügung gestellt werden. CHEVAL ermöglicht es, eine identische Untersuchungsumgebung mit gleichen Testkollektionen und Testaufgaben für unterschiedliche Systeme zu erzeugen. Bei der Vorbereitung einer Evaluation wird der Testadministrator auf zweierlei Arten unterstützt. Zum einen erhält er aufgrund des bereits beschriebenen Entscheidungsbaums Vorschläge, welche Methoden für die Evaluation des zu untersuchenden Systems geeignet sind. Weiterhin erleichtert CHEVAL die Analyse und den Vergleich der Testergebnisse verschiedener Systeme, da die Testresultate stets identisch strukturiert dargestellt werden. Durch die Verwendung von CHEVAL, entfällt für betroffene Unternehmen die Entwicklung

und Planung eigener Evaluationskonzepte, wodurch der Evaluationsprozess wesentlich verkürzt und Kosteneinsparungen realisiert werden können.

4.3.2 Wirtschaftlichkeitsanalyse von Informationsdiensten

Mit Hilfe von CHEVAL werden weiterhin Wirtschaftlichkeitsanalysen von Informationsdiensten ermöglicht. Ein Beispiel aus diesem Bereich ist der ursprünglich von der Massachusetts Library Association entwickelte „Library Use Value Calculator“. Dieser bietet die Möglichkeit, mittels eines einfachen Formulars den Wert von Bibliotheksleistungen zu quantifizieren. In CHEVAL werden verschiedene solcher Tools bereitgestellt werden. Den Anwendern soll die Möglichkeit geboten werden, diese nach dem Baukastenprinzip an ihre eigenen Bedürfnisse anpassen zu können. Dabei erhalten die Nutzer Unterstützung über die Wissensbasis von CHEVAL. Dort werden verschiedenen Masse und Methoden zur Bewertung der Wirtschaftlichkeit von Informationsprodukten dargestellt. Diese Informationen können von Anwendern genutzt werden, um sich ein für ihren Untersuchungszweck massgeschneidertes Bewertungstool zu erstellen. Auf diese Weise trägt CHEVAL zu einem wirtschaftlicheren Umgang mit der Ressource Information bei.

5 CHEVAL – ein Fazit

Die im vorangegangenen Kapitel dargestellten Anwendungsfälle verdeutlichen, dass den Nutzern mit CHEVAL ein breites Spektrum an Anwendungsmöglichkeiten geboten werden soll. Der Ansatz verschiedener Methoden aus den Bereichen der Retrievaleffektivitätsmessung, dem Usability-Testing, der Nutzerbefragung und der Wirtschaftlichkeitsanalyse in einer Evaluationsumgebung zu integrieren, ist in dieser Form neu.

Ein entscheidender Faktor für den Erfolg von CHEVAL ist die Frage, ob am Markt eine entsprechende Nachfrage nach den geplanten Leistungen besteht. CHEVAL soll Unterstützung für unterschiedlichste Arten von Evaluationen bieten, wie z.B. Usability Tests, Retrieval-Effektivitätsmessungen, Wirtschaftlichkeitsmessungen oder auch Kundenzufriedenheitsanalysen. Ähnlich integrierte Angebote sind derzeit auf dem Markt nicht verfügbar. So wurde in einer Marktuntersuchung festgestellt, dass kein Anbieter existiert, der Usability-Tests von IRS in Kombination mit einer Retrieval-Effektivitätsmessung anbietet. Weiterhin hat auch eine in Zusammenarbeit mit mehreren Unternehmen und einer weiteren Hochschule durchgeführte Studie, bei der die Suchfunktionen der Webseiten grosser, schweizerischer Unternehmen getestet wurden, ergeben, dass ein Markt für eine solche Dienstleistung vorhanden ist, zumal die meisten der untersuchten Webseiten teils erhebliche Mängel aufwiesen (Braschler et al. 2006). Zusätzliches Potenzial ist darin zu sehen, dass mit Hilfe von CHEVAL auch Benchmarkingstudien durchgeführt werden können, wobei entweder ein Vergleich mit den Systemen einer bestimmten Branche oder ein Vergleich mit Systemen der selben Art möglich ist.

CHEVAL verfolgt das Ziel, ein Rahmenwerk für die standardisierte und systematische Evaluation solcher Systeme bereit zu stellen. Auf Grund der modularen Struktur der Testumgebung und des modularen Aufbaus wird das System einfach erweiterbar sein. Dadurch wird ermöglicht, neue Entwicklungen im Bereich von IRS, wie beispielsweise eine „Social Search“ auf Basis von Web 2.0 Komponenten und andere Trends zu berücksichtigen und entsprechende Evaluationskomponenten in das System integrieren zu können.

Durch das breite Leistungsspektrum von CHEVAL soll das System allen Unterstützung bieten, die sich in irgendeiner Form mit Evaluationen beschäftigen, sei es in Form von Kunden- und Mitarbeiterbefragen, Usability-Tests oder auch der Wirtschaftlichkeitsanalyse von Informationsdiensten. Dadurch sollen Evaluationen schneller und effizienter durchgeführt werden können, wodurch sich die Kosten von Evaluationen senken lassen, bei gleichzeitiger Qualitätserhöhung durch eine Standardisierung der verwendeten Methoden und Instrumente.

Literatur

Arnold, C. und Wolff, C. (2005) "Evaluierung von Visualisierungsverfahren bei der webbasierten Suche.", *Knowledge eXtended. Die Zusammenarbeit von Wissenschaftlern, Bibliothekaren und IT-Spezialisten*, Schriften des Forschungszentrums Jülich, Reihe Bibliothek, Band 14, 2005, 275-286.
Bauer, L., Herget, J. und Hierl, S. (2007) "Conceptual challenges for the evaluation of digital repositories with multiple access options", *Conference on Semantic Web & Digital Libraries (ICSD)*, Bangalore (India), Feb. 2007.

- Bekavac, B., Herget, J., Hierl, S. und Öttl, S. (2007) "Visualisierungskomponenten bei Web-basierten Suchmaschinen: Methoden, Kriterien und ein Marktüberblick." *IWP - Information Wissenschaft & Praxis*, 58 (2007) 3, 149-158.
- Braschler, M., Herget, J., Pfister, J., Schäuble, P., Steinbach, M. und Stuker, J. (2007) „Kommunikation im Internet: Die Suchfunktion im Praxistest.“ *IWP - Information Wissenschaft & Praxis*, 58 (2007) 3, 159-168.
- Cross-Language Evaluation Forum (CLEF) [Online], <http://www.clef-campaign.org/>.
- Google [Online], <http://www.google.de>.
- Grokker [Online], <http://live.grokker.com>.
- Harman, D. (1993) "Overview of TREC-1", *Proceedings of the workshop on Human Language Technology*, Princeton, New Jersey, 61-65.
- Hierl, S. (2007) "Bezugsrahmen für die Evaluation von Information Retrieval Systemen mit Visualisierungskomponenten", *B.I.T. Online*, 2/2007.
- Koshman, S. (2005) "Testing user interaction with a prototype visualization-based information retrieval system", *Journal of the American Society for Information Science and Technology*, 56(8) 2005, 824-833.
- Krcmar, H. (2005) "Informationsmanagement", Berlin, Heidelberg .
- Lewandowski, D. und Höchstötter, N. (2007) „Qualitätsmessung bei Suchmaschinen – System- und nutzerbezogene Evaluationsmaße“, *Informatik-Spektrum*, 30 (3) 2007 159-169.
- Library Use Value Calculator [Online], <http://www.vermontlibraries.org/value.html>.
- Massachusetts Library Association [Online], <http://www.masslib.org/>.
- MSN [Online], <http://de.msn.com/>.
- Quintura [Online], <http://www.quintura.com/>.
- Reiterer, H. (2004) "Visuelle Recherchesysteme zur Unterstützung der Wissensverarbeitung", Hamwöhner, R., Rittberger, M., Semar, W. (Eds.): *Wissen in Aktion. Der Primat der Pragmatik als Motto der Konstanzer Informationswissenschaft. Festschrift für Rainer Kuhlen*. Constance, 1–21.
- Reiterer, H., Tullius, G., und Mann, T. M. (2005) „INSYDER: a content-based visual-information-seeking system for the Web“, *International Journal on Digital Libraries*, Volume 5, Issue 1, Mar. 2005, 25 - 41.
- Schoop, E. (2004) "Information in der Betriebswirtschaft: ein neuer Produktionsfaktor?" in: Kuhlen, R., Seeger, T. und Strauch, D. (Hrsg.): *Grundlagen der praktischen Information und Dokumentation*. München, 717-721.
- Sebrechts, M., Vasilakis, J., Miller, M., Cugini, J. und Laskowski, S. (1999) "Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces" In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 3-10.
- Text REtrieval Conference (TREC) [Online], <http://trec.nist.gov/>.
- Ujiko [Online], <http://www.ujiko.com/>.
- Vaughan, L. (2004) "New measurements for search engine evaluation proposed and tested", *Information Processing and Management* 40 (2004) 677–691.
- Veerasamy, A. und Belkin, N. J. (1996) "Evaluation of a tool for visualization of information retrieval results", *Proceedings of the 19th Annual international ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '96*, ACM Press, New York, NY, 85-92.
- Voorhees, E. and Harman, D. (2005) "The Text Retrieval Conference", Voorhees, E and Harman, D. (Eds.): *TREC - Experiment and Evaluation in Information Retrieval*, The MIT Press, Cambridge.
- Yahoo [Online], <http://de.search.yahoo.com>.
- Zwol, R. Van und Oostendorp, H. Van (2004) "Google's 'I'm feeling lucky', Truly a Gamble?", Zhou, X. et al. (Eds.): *Web Information Systems - WISE 2004, Proceedings of the 5th International Conference on Web Information Systems Engineering*, Brisbane, Australia, 378-390.