

**HTW** Chur  
Hochschule für Technik und Wirtschaft

Schweizerisches Institut für  
Informationswissenschaft

# **Webinfolab / CHEVAL: Evaluationslabor für Information Retrieval Systeme mit semantischen und visuellen Komponenten**

Vortrag anlässlich der ODOK 2007 (20. September 2007, Graz)

Joachim Pfister

Schweizerisches Institut für Informationswissenschaft (SII)  
Hochschule für Technik und Wirtschaft, HTW Chur, Schweiz

# Agenda

1. Evaluierung von Information Retrieval Systemen (IRS)
2. Motivation
3. Webinfolab / CHEVAL
  - Ziele
  - Evaluationsprozess
  - Architektur
4. Anwendungsszenario
5. Fazit und Ausblick

# Evaluierung von IRS

## **Ausgangspunkt:** Cranfield-Experimente (1962)

- gemeinsame Testkollektionen (Dokumente und Anfragen)
- vergleichende Evaluation

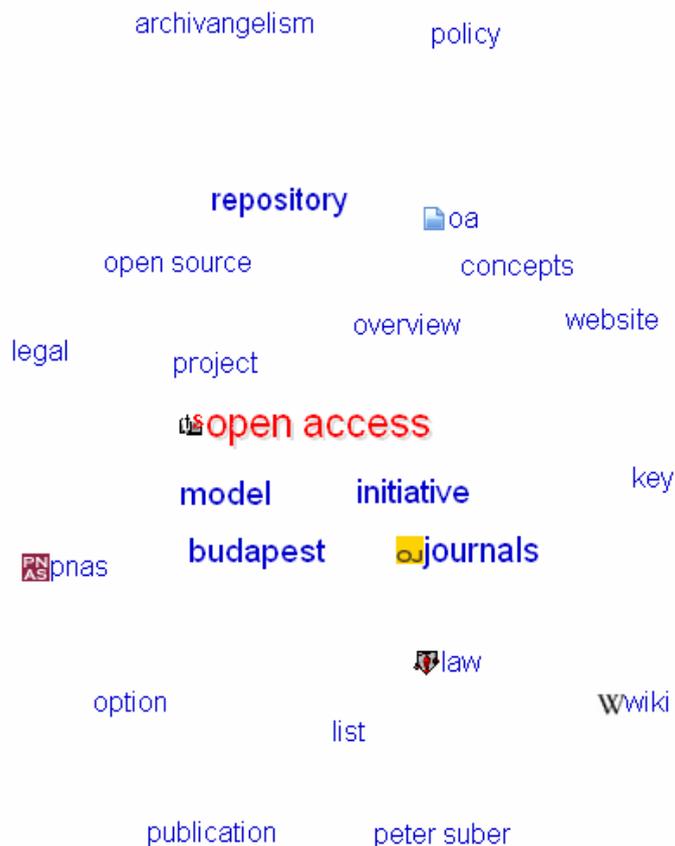
(vgl. Voorhees und Harmann, 2005)

## **Bekannte Evaluierungs-Initiativen**

- TREC (Text REtrieval Conferences)
- CLEF (Cross Language Evaluation Forum)
- INEX (INitiative for the Evaluation of XML retrieval)

## **Weiterentwicklungen der Technologien**

- Visualisierung von Suchergebnissen
- Semantische Konzepte
- Web2.0 Konzepte



1. [Open access - Wikipedia, the free encyclopedia](#)  
 For other uses, see Open access (disambiguation)... This provided a de...  
 has a growing list of signatories...  
[http://en.wikipedia.org/wiki/Open\\_access](http://en.wikipedia.org/wiki/Open_access), 99Kb
2. [Budapest Open Access Initiative](#)  
 The Budapest **Open Access** Initiative: an international effort to make re...  
 academic fields freely available on the internet.  
<http://www.soros.org/openaccess/>, 11Kb
3. [Open Access](#)  
 Open Access receives Certificate of Appreciation and Giraffe Award from...  
 at the ASSET 2006 Educational Technologies Conference. ...  
<http://www.openaccessinc.com/>, 10Kb
4. [open access](#)  
<http://www.openaccess.tw/>, 4Kb
5. [Peter Suber, Open Access News](#)  
 ... Open Medicine and **open access** , Canadian Medical Association J...  
 Request Button, **Open Access** Archivangelism, September 15, 2007. ...  
<http://www.earlham.edu/~peters/fos/fosblog.html>, 247Kb
6. [Open Access Website](#)  
 Help. Register. Log in. Welcome to the Open Access Website for Land...  
**access** land. and guidance for owners and managers of access land. ...  
<http://www.openaccess.gov.uk/wps/portal>, 34Kb
7. [Peter Suber, Open Access Overview \(definition, introduction\)](#)  
 Focusing on **open access** to peer - reviewed research articles and... w...  
 access movement... Commons Through **Open Access** : ...  
<http://www.earlham.edu/~peters/fos/overview.htm>, 47Kb
8. [Welcome to Open Access](#)  
 Offers advice on access rights under the Countryside and Rights of Wa...  
 features public walking and restriction maps, workshops for land manag...  
 resources for walkers.  
<http://www.openaccess.gov.uk/>, 10Kb



Selected Sources [2 of 3] [Add/Remove](#)  
Yahoo!, Wikipedia

"Open Access"

**GROK** Search Options

Outline View **Map View** 138 total results

ZOOM BACK **TOP**

Working List  
0 items in your list  
[View your list](#)

[Email Map...](#)  
[Export Map...](#)

Search within the map:

keyword   Exclude

date   
06/2007 to most recent

source  
All sources >

domain  
All domains >

**Hide Tools**

Color



<b>Title</b>	BioMed Central   about us   Open access charter
Date	16.09.2007
Rank	19
Source	Yahoo!

# Motivation

- Evaluation von IRS basiert auf Standardmassen (Recall, Precision, Kombinationsmasse [z.B. F-Measure])
- Usability eines IRS wird meist separat bewertet (z.B. Screencapturing, Guidelines, Tagebuchstudien)
- Einflussfaktoren für IRS-Evaluationen:
  - Komplexe Zusammenhänge zwischen
    - Ergebnispräsentation (Benutzungsoberfläche/GUI)
    - Interaktionsmöglichkeiten
    - Semantischen Technologien
  - Gewöhnungseffekte (De-Facto Standard Google, Yahoo oder MSN)



Eine umfassende Evaluation eines IRS (mit Schwerpunkt aus Nutzersicht) muss sowohl Standardmasse als auch Usability-Aspekte kombinieren!

# CHEVAL – Ziele und Vision

- Webinfolab → CHEVAL (CHur EVAluation Laboratory)  
<http://www.cheval-lab.ch> (bislang: WIKI)
- CHEVAL als Framework, dass
  - verschiedene Evaluationsansätze und –methoden bereitstellt,
  - verschiedene Arten der Evaluation von IRS-Systemen unterstützt.
- CHEVAL bietet eine Infrastruktur zum
  - Durchführen von Evaluationen
  - systemunterstützten Auswählen geeigneter Methoden
  - Verwalten von Testkollektionen
  - Überwachen des Verlaufs eines Evaluationsprojekts
  - Speichern der Ergebnisse, auch für Benchmarking
  - Generieren von Berichten

# Tool mit relevanter Teilfunktionalität (bspw. als Methode in CHEVAL zu integrieren)

Retrieval Performance Test mittels des Tools "TERS" als Beispiel

TERS - Retrieval Performance Interface

http://contentlab.cs.uu.nl/~testbed/demo/rpe/index

## TERS - Retrieval Performance Experiment

Introduction --> Login --> Topic selection --> **Relevance assessments** --> Finish [help]

[Gathering relevance assessments for topic 2 by guest]

### Topic execution

Topic: **2. Decompression illness**

**Description:** What is decompression illness? 1.

**Narrative:** *The "bends" is today readily associated with SCUBA diving. It is, in fact, an old-fashioned term used originally to describe the appearance of workers returning from 'caissons' during the construction of bridges in the 19th century. The air inside these underwater enclosures was pressurised (i.e. hyperbaric - greater than normal pressure) to counteract the weight of the surrounding water. Following their shifts, some men would return to the surface suffering joint pain that made it difficult for them to stand straight. Their appearance was similar to the 'Grecian bend' adopted by fashionable women of the time - hence the name. Many workers died or suffered permanent disability because of "Caisson disease", as the condition became known. The connection between the workers' return to the surface and their symptoms led to the introduction of surface based re-compression chambers to treat the afflicted.* 2.

Search phrase: decompression illness

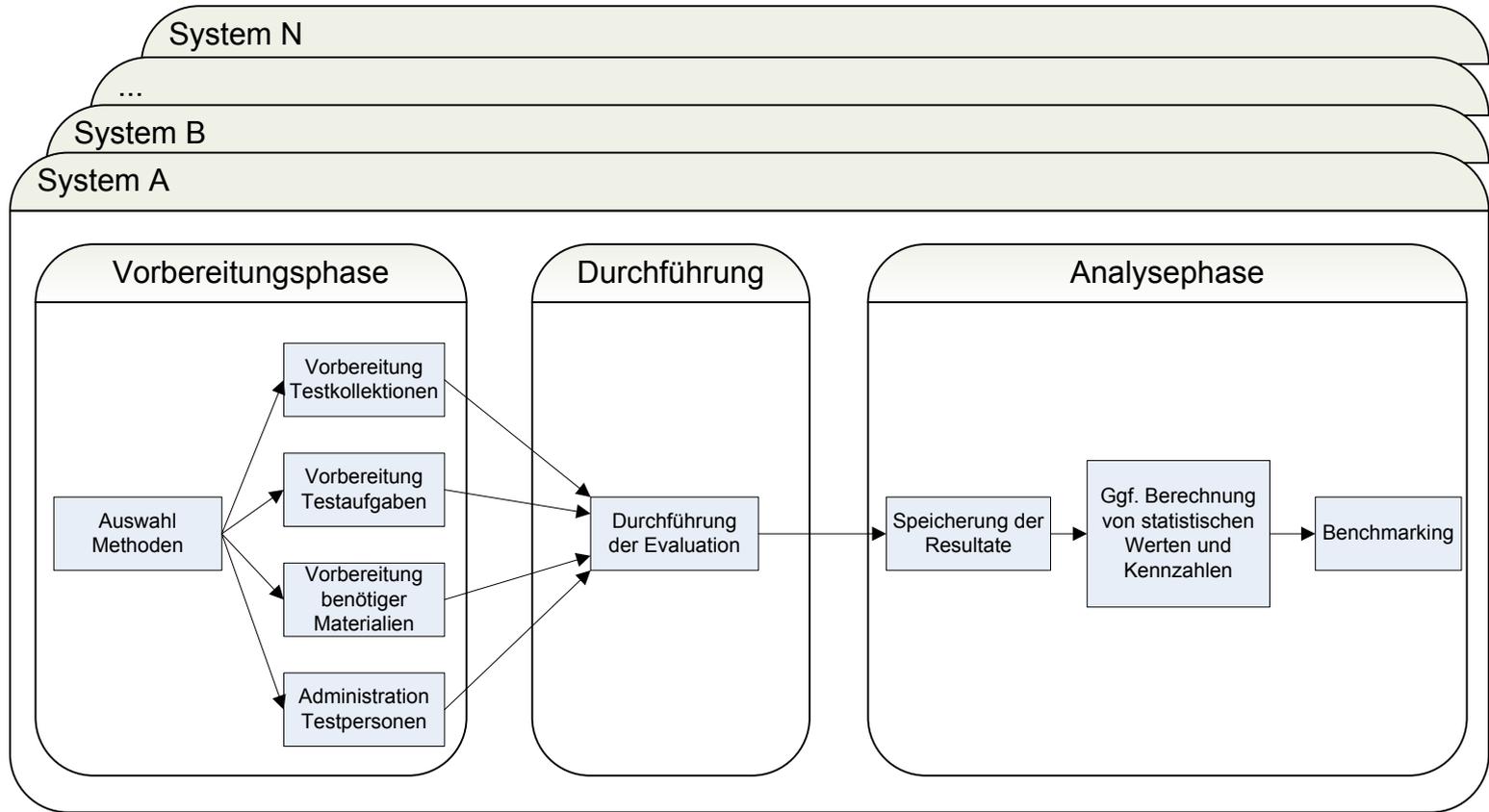
### Results for topic 2. Decompression illness

Relevance		Uri
--	- 0 + ++	
4.		http://en.wikipedia.org/wiki/Decompression_illness
		http://www.mindwellness.com/decompression_illness.htm
		http://www.divegirl.com/bent.html
		http://www.diversalertnetwork.org/medical/articles/article.asp?articleid=65
		http://www.emedicinehealth.com/articles/27678-1.asp

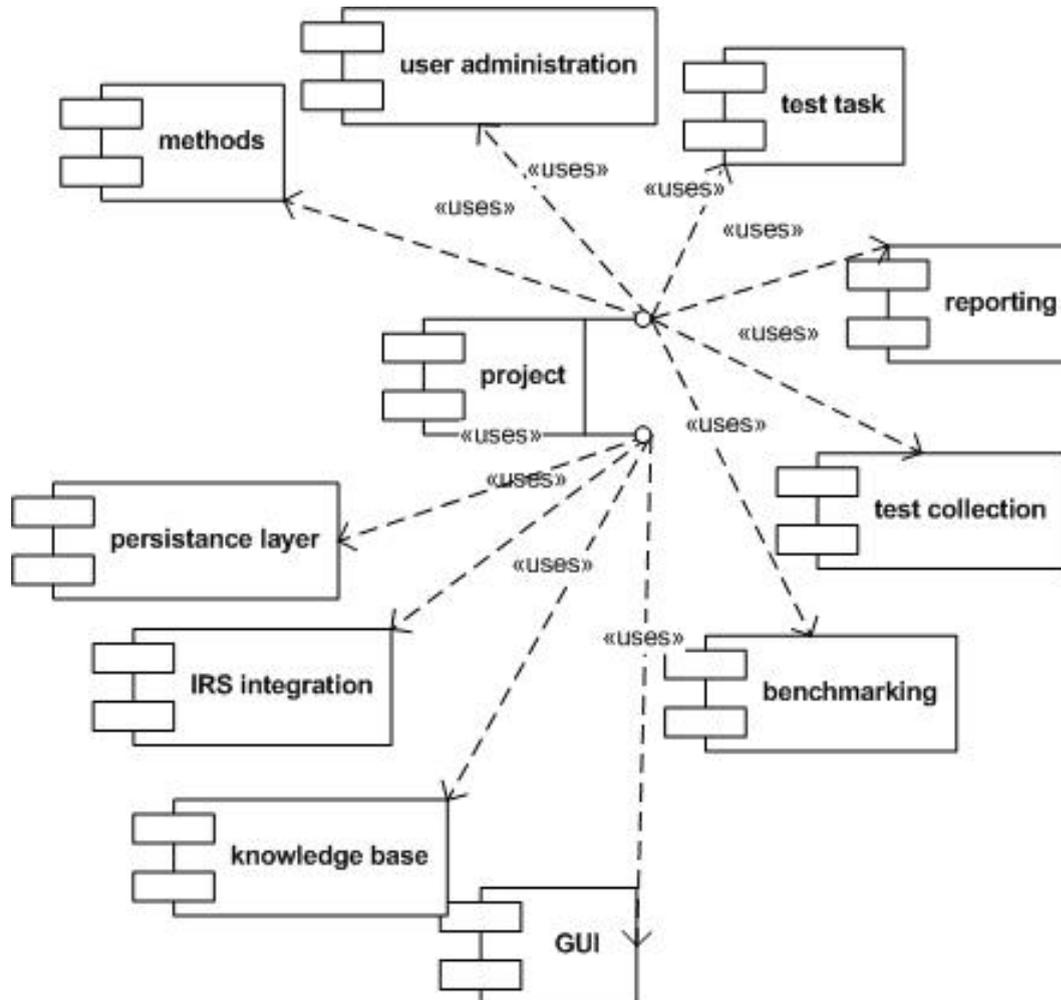
Done

(van Zwol und Ossstendorp, 2004)

# Phasen einer Evaluation – unterstützt durch CHEVAL

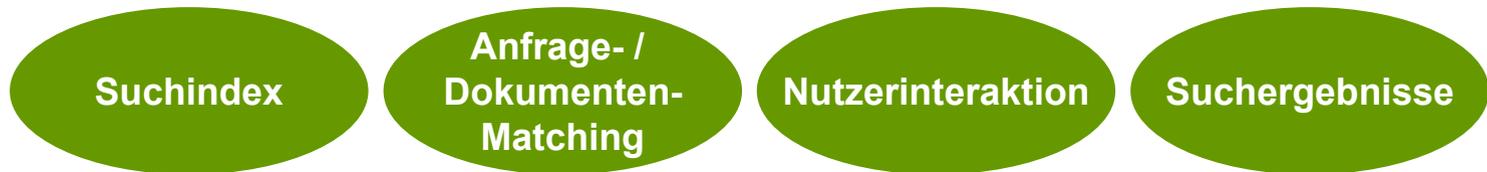


# CHEVAL - Architektur



# CHEVAL – Anwendungsszenario (I.)

- Einsatzszenario für CHEVAL:  
Evaluation der Suchfunktion von Unternehmens-Websites
- Zwei Studien (Herbst 2006: CH, Frühjahr 2007: D), bei der mit weiteren Hochschul- und Wirtschaftspartnern die Suchfunktionalität von Unternehmenswebsites und Websites Öffentlicher Einrichtungen untersucht wurde. (Braschler et al. 2006 und 2007)



- Bisher: Medien- und Anwendungsbruch
  - Aufgabenbeschreibung in Papierform
  - Erfassung in Excel-Tabelle (Zwischenergebnisse, Resultate)
  - Durchführung der Tests im Browser

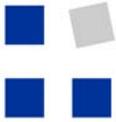
# CHEVAL – Anwendungsszenario (II.)

Mit Hilfe von CHEVAL:

- Verknüpfung von Aufgabenbeschreibung und Möglichkeiten der Ergebniserfassung auf einer Web-Seite
- Kein Medien- oder Anwendungsbruch
- Ermöglicht das Testen auf Korrektheit und Vollständigkeit (höhere Datenqualität)
- Überblick über den aktuellen Status eines Testprojekts (wie viele Tester haben wie viel schon bearbeitet?)
- Erstellen einer Benchmarking-Datenbank, z.B. im Rahmen
  - einer vergleichenden Evaluation (verschiedene Anbieter eines IRS),
  - der Implementierung eines IRS, um die Konfiguration mit den „optimalen“ Parametern zu ermitteln.

# Fazit und Ausblick

- **CHEVAL:**  
Framework zum Durchführen und Auswerten von Evaluationen
- Konkrete Programmierung im Q1 2008 geplant
- Inkrementelles Vorgehen bei der Implementierung:  
Nach zunehmender Komplexität werden Funktionalitäten ergänzt.
- Geplante Evaluations-Dienste (in der Reihenfolge ihrer Umsetzung):
  - Website Search Assessment
  - Kundenzufriedenheitsanalyse
  - IRS Beratungsdienstleistungen
  - IRS Effizienzmessungen (Retrieval Performance Test)
  - Usability Tests
  - Evaluation von IRS mit visuellen oder semantischen Konzepten
- Bereitstellung der Software als Open Source Software (geplant)
  - Beitrag für die wissenschaftliche Community
  - Weiterentwicklung durch die Community



# Fragen?

# Vielen Dank für Ihre Aufmerksamkeit.

## **Ausführlicher Artikel:**

Bauer, L.; Herget, J.; Hierl, S.; Pfister, J.; Weinhold, T. (2007): Cheval: Das Churer Evaluationskonzept für Information, Systeme und Nutzer - eine Toolbox im Entstehen. Proceedings der DGI-Tagung 2007, 10.-12. Oktober 2007 in Frankfurt am Main, Seiten 285-292.

Weinhold, T.; Bauer, L.; Herget, J.; Hierl, S.; Pfister, J. (2007): CHEVAL: Chur Evaluation Laboratory. Proceedings of the European Conference on Information Management and Evaluation 2007 (ECIME 2007), September 20-21, 2007, Montpellier, France. Pages 393-400.

# Quellen

Braschler, Martin; Herget, Josef; Pfister, Joachim; Schäuble, Peter; Steinbach, Markus; Stuker, Jürg (2006): Evaluation der Suchfunktion von Schweizer Unternehmens-Websites. In: Herget, J; Hierl, S. (Hrsg.): Churer Schriften zur Informationswissenschaft – Schrift 12, 2006

Braschler, Martin; Herget, Josef; Pfister, Joachim; Schäuble, Peter; Steinbach, Markus; Stuker, Jürg: Kommunikation im Internet: Die Suchfunktion im Praxistest. In: IWP - Information Wissenschaft & Praxis 58 (2007) 3, S. 159-168

Voorhees , Ellen M. und Harman, Donna K. (2005): The Text REtrieval Conference. In: TREC. Experiment and Evaluation in Information Retrieval. MIT Press : Cambridge, Mass. <http://mitpress.mit.edu/books/chapters/0262220733chap1.pdf>, verifiziert am 18.09.2007, 11:03 Uhr MEZ

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.