



Die SEMPRIA[®]-Suchmaschine:

Mehr erreichen als mit Stichwortsuche

Präsentation für den Runden Tisch Is

Köln, 2012-08-20

Dr. Sven Hartrumpf, s.hartrumpf@sempria.de, 0211/566693-57
SEMPRIA GmbH, Grafenberger Allee 277–287, 40237 Düsseldorf
<http://www.sempria.de/>

SEMPRIA GmbH, 2012-08-20

Problem des Kunden und Lösung von SEMPRIA



1. Problem

Archivbesitzer (Verlage, Rundfunkhäuser, Web-Sites, ...):
haben **Probleme**, ihre Archive voll und effizient zu **erschließen**

2. Lösung

Suchmaschine SEMPRIA[®]-Suchmaschine hilft durch

- ▶ **höhere Vollständigkeit** von Suchergebnissen
- ▶ **höhere Genauigkeit** von Suchergebnissen
- ▶ **Fragefunktionalität** (Möglichkeit gezielter Fragen)

3. Wie?

eigene Technologie: **automatisches Sprachverstehen für Deutsch**

4. Alleinstellungsmerkmal der Lösung

durch Einsatz von Sprachverstehen deutliche Leistungssteigerungen

- ➔ Nutzer (Leser, Redakteure, Analysten, ...) **sparen Recherchezeit**
- ➔ Nutzer **finden öfter das einzige relevante Dokument**

SEMPRIA GmbH, 2012-08-20



Ziel einer Site-Search-Lösung:

wertvolle **Informations-Nuggets** in einer Goldgrube (= Site) effizient finden

10 Ergebnisseiten mit jeweils 10 Treffern, **richtige Treffer**

Suchanfrage *Lincolns Tod* ⇒

Seite 1				☺					
Seite 2					☺			☺	
Seite 3									
Seite 4		☺							
Seite 5						☺			
Seite 6							☺		
Seite 7	☺								
Seite 8									
Seite 9									☺
Seite 10							☺		

Live-Test:

Tod ... (unverstandener Kontext) ... **Lincoln Continental** =? **Lincolns Tod**

Vergleich auf Ebene von **Wortsäcken** (*bag-of-words*)

SEMPRIA GmbH, 2012-08-20



Suchanfrage *Lincolns Tod*

⇓ hunderte Umformulierungen durch automatisches Sprachverstehen

25 Ergebnisse für Umformulierung 1

25 Ergebnisse für Umformulierung 2

		☺	☺						
	☺								
			☺						
☺									

...

⇓ weitere Filterung durch Abgleich der Bedeutungen

3 Ergebnisseiten für Nutzer

Seite 1	☺	☺	☺		☺	☺		☺	
Seite 2	☺		☺		☺			☺	
Seite 3		☺	☺		☺	☺	☺		☺

⇓ Optional bei Fragefunktionalität (*Wann starb Lincoln?*)

☺ **am 14.04.1865** (knapper Beleg für die Antwort)

Live-Test:

Lincoln¹ ... (verstandener Kontext) ... **fiel er**¹ ... **zum Opfer** = **Lincolns Tod**

Systemintern enthalten beide (Anfrage und Dokument):

sterben.1.1 –EXPERIENCER→ Mensch –NAME→ Lincoln

Vergleich auf Ebene von **Bedeutungsstrukturen**

SEMPRIA GmbH, 2012-08-20



Nutzeranfrage:

Import von Computern

Zusätzliche richtige Treffer gegenüber Stichwortsuche (→ Vollständigkeit):

Einfuhr an Rechnern stieg ...

importierte 2010 ... Laptops aus ...

führt im nächsten Jahr ... Workstations aus

Vermiedene falsche Treffer der Stichwortsuche

(→ Genauigkeit durch semantische Filter):

um den Import mit Computern besser zu überwachen



Nutzeranfrage:

Wo befindet sich der schnellste Rechner ?

Gefundene Belegstelle:

Schnellster Computer Deutschlands steht in Stuttgart .

Ermittelte Antwort:

in Stuttgart



- ▶ **Suchfunktion (im Web oder in-house) für Textarchive von:**
 - Zeitungen, Zeitschriften
 - Radio, Fernsehen
 - Enzyklopädien, Akten, Dokumentation . . .
- ▶ **Alternatives Suchmodul für Content-Management-Systeme**
- ▶ **Unternehmensweite Suche (*enterprise search*)**
- ▶ **Archivierungsunterstützung**
 - Verschlagwortung
 - Verlinkung
 - Duplikatserkennung . . .
- ▶ **Sentimentanalyse (*opinion mining, issue management*)**
in Vorbereitung

SEMPRIA GmbH, 2012-08-20

Kosten sparen, Qualität steigern: Site-Search



Einsatzfeld Site-Search für Webseiten-Besucher:

- ▶ Akzeptanz der Site-Search steigern
- ▶ Abwanderung zur Google-Websuche stoppen
- ▶ höhere Verweildauer auf eigenen Seiten erzielen

Einsatzfeld Site-Search für Bezahlkunden:

mehr und passendere Artikel für Online-Verkauf anbieten

Grobrechnung (Website mit 1 Mio. Seitenaufrufen/Monat, Kaufanteil: 0,5%):

1.000.000 Aufrufe/Monat * 0,5% Kaufanteil * 50% Kaufanteilsverbesserung *

0,50 €/Artikel * 12 Monate = **15.000 € p.a.**

(Ihre Parameter in dieser Farbe)

SEMPRIA GmbH, 2012-08-20



Traditionelle Suchmaschinen:

Flexion, Derivation, Komposita (teils unvollständig)

Nutzung von Metadaten

facettierte Suche für einige Namensklassen

SEMPRIA® (zusätzlich):

Mehrwortausdrücke → *Rio de Janeiro* stets als Einheit

Zahlen, Maße → *10 km* ↔ *10000 Meter* ↔ *10.000 m ...*

Idiome, Metonymie * → *Handtuch werfen* ↔ *aufgeben*

Funktionsverbgefüge * → *Antrag stellen* ↔ *beantragen*

Mehrdeutigkeiten (lexikalische, strukturelle) * → *Lincoln: Mensch, Auto ...*

Zeitangaben (absolute und relative)

→ *vorgestern (oder am Mittwoch ...)* + Metadaten ↔ *31.07.2011*

facettierte Suche mittels voller Semantik

* abhängig vom verfügbaren Hintergrundwissen

SEMPRIA GmbH, 2012-08-20



Bezüge von Pronomen * → *Die X¹ kritisierte den Y. Sie¹ antwortete ...*

Beziehungen zwischen Objekten → *Was hat Firma X mit Kriegswaffen zu tun?*

Rolle von Beteiligten in Ereignissen

→ *Kauf der Dresdner Bank vs. Kauf durch Dresdner Bank*

Bedeutung aus mehreren Sätzen/Dokumenten → *Wer exportiert seltene Erden?*

→ *Dokument 1: seltene Erde Neodym + Dokument 2: X führt Neodym aus*

semantische Suchvorschläge aus den Dokumenten

Fehlerrobustheit (Rechtschreibung, Zusammenschreibung)

→ *Fussball WM* ↔ *Fußball-WM*

* abhängig vom verfügbaren Hintergrundwissen

SEMPRIA GmbH, 2012-08-20



SEMPRIA GmbH
Grafenberger Allee 277–287
40237 Düsseldorf

Telefon: 0211/566693-57
Fax: 0211/566693-58
Web: <http://www.sempria.de/>
E-Mail: info@sempria.de

Geschäftsführer: Dr. Sven Hartrumpf
Handelsregister: Amtsgericht Düsseldorf, HRB 62168
UStID-Nr: DE268248179