

Produktdatenblatt der bedeutungsorientierten Suchmaschine SEMPRIA-Search

SEMPRIA GmbH, Grafenberger Allee 277-287, 40237 Düsseldorf, <https://www.sempria.de/>,
Stand: 2018-06-01

Anwendungsbereiche

Mit der SEMPRIA-Suchmaschine werden Dokumentensammlungen und Archive auf neuartige Weise besser und intelligenter erschlossen als bei traditionellen Suchmaschinen. Durch Erkennen der vollen Bedeutung von Suchanfragen und Dokumenten liefert diese bedeutungsorientierte Suche qualifiziertere Antworten. Einsatzmöglichkeiten als:

- Suchmodul für Content-Management-Systeme (CMS), Dokumenten-Management-Systeme (DMS), Online-Shops und Wikis
- Suchfunktion für Archive von Zeitungen, Zeitschriften, Radio, Fernsehen
- Unternehmensweite Suche (Enterprise Search)

Ihre Vorteile

- SEMPRIA-Search erhöht den Anteil richtiger Ergebnisse → steigert die Zufriedenheit der Suchenden
- SEMPRIA-Search findet auf intelligentem Weg Dokumente, die eine traditionelle Suchmaschine nicht finden kann → maximiert das Verwertungspotenzial Ihrer Texte
- SEMPRIA-Search bietet erstmals eine bedeutungsorientierte Suche für deutschsprachige Archive → sichert Ihren Innovationsvorsprung

Produktmerkmale

Die Formulierungen aus den gestellten Suchanfragen werden intelligent mit den Formulierungen aus Ihren Dokumenten verknüpft. Es folgen einige Highlights und Beispiele für semantische Suche, wie sie mit SEMPRIA-Search möglich sind:

- Nominalisierungen (*Erschließung* ist eine Nominalisierung des Verbs *erschließen*): *Erschließung von Information* (Suchanfrage) findet *Information erschließen* (Dokumentenstelle); hier und in den meisten folgenden Beispiele sind die Rolle von Suchanfrage und Dokumentenstelle vertauschbar.
- Begriffshierarchien mit Oberbegriffen, Unterbegriffen und synonymen Begriffen (Synonyme)
- Funktionsverbgefüge: *beantragen* findet *Antrag stellen*
- Mehrwortausdrücke: *seltene Erden* findet (zusammen mit Begriffshierarchien, s.o.) *Neodym*
- Komposita (Zusammensetzungen): *Briefkastenhersteller* findet *Briefkästen herstellt*. (Abweichende Komposita, die keine regelhafte Bedeutung haben, werden erkannt, so dass hier *Briefkastenfirma* als Suchtreffer richtigerweise verworfen wird.)



- idiomatische Ausdrücke: *CEO gibt auf* findet (zusammen mit Begriffshierarchien) *Geschäftsführer wirft das Handtuch*
- Ausnutzung von Schlussfolgerungen: *Akkuimport nach Deutschland* findet *exportiert Akkus nach Deutschland*
- Robustheit gegen Schreibfehler: *Feehler* vs. *Fehler*
- Robustheit gegen verschiedene Orthographien: *Abguss* vs. *Abguß*, *aufwändig* vs. *aufwendig*, *Fotograf* vs. *Photograph*
- gezielte Beantwortung von Fragen: *Wann wurde die Filiale in Bayern geöffnet?* findet die genauen Daten (vorausgesetzt, es gibt solche Informationen)
- Indexierung von Dokumenten verschiedenster Formate (PDF, Office, LibreOffice, HTML etc.)
- Indexierung von Scans und Fotos durch Anbindung an eine präzise Zeichenerkennung (OCR) eines Marktführers
- Indexierung von Audiodokumenten und Videodokumenten durch eine genaue Spracherkennung mit eigener intelligenter Integration in das Sprachverstehen der Suchmaschine
- vollständige und clevere Integration der Metadaten des CMS/DMS in das Such-Interface.

Das manuell gepflegte Wissen, das wir u.a. durch Methoden der künstlichen Intelligenz (KI) und des maschinellen Lernens (ML, machine learning) erweitern, umfasst mittlerweile mehrere Millionen Regeln und Fakten. Täglich wird dieses Wissen von Linguisten und Wissensingenieuren weiter ausgebaut, denn nur so sind die obigen Produktfeatures überhaupt möglich.

Technische Daten

65.000	semantisch umfassende Einträge im Computerlexikon
500.000	Eigennamen in 50 Klassen
1.500.000	Analysen von Komposita (zusammengesetzte Wörter)
3.000.000	Fakten und Regeln im Hintergrundwissen für die KI-Komponenten
300.000	Zeilen Programmcode in KI-Programmiersprachen

Praktischer Einsatz

- als Software-as-a-Service (SaaS) oder als Server-Lösung (*on-premise*)
- Einbindung in bekannte CMS und DMS durch kleinen Codeblock (JavaScript o.ä.)
- Update des Suchindexes in vereinbarten Abständen (täglich, wöchentlich, monatlich ...)
- Auswertungen der Suchanfragen (Suchreport): regelmäßig, anpassbar
- Auswertungen zum Sprachgebrauch in den Dokumenten (Sprachreport): regelmäßig zu Tippfehlern, Grammatikfehlern ...

Preisgestaltung

Preise für SEMPRIA-Search richten sich nach drei Parametern: Größe der Dokumenten-Sammlung, Volumen der monatlichen Änderungen und Anzahl der Suchanfragen pro Stunde. Gern schicken wir Ihnen anhand Ihrer Parameter-Angaben eine übersichtliche Preismatrix oder ein individuelles Angebot. SEMPRIA-Search kostet bei kleineren Systemen im Monat typischerweise 40 EUR.

