

## Kurzbeschreibung ht://Dig

Bezeichnung: ht://Dig

Funktion: Volltextsuch-Programm

Beschreibung: 1. Index-Erzeugung:

ht://Dig indiziert Web-Sites via http , wobei die Steuerung über einen Konfigurationsfile erfolgt. Indiziert werden können HTML und Text Dateien. Für PDF bzw. MS Office Dateien können externe Parser eingebunden werden.

2. Suche:

Zur Suche wird ht://Dig über die CGI-Schnittstelle eines Web-Servers betrieben. Die WWW-Benutzeroberfläche ist kann über HTML templates relativ einfach angepasst werden. Die Abfrage kann Boolesche Operatoren und Klammern enthalten. Die Tiefe der Suche kann eingeschränkt werden. ht://Dig unterstützt die METATAG keywords suche.

3: Ergebnis Ausgabe

Die Ausgabe kann kann über HTML templates relativ einfach angepasst werden.

Anwendung im ITZ KA: Mittels ht://Dig wird der Intranetserver der LfU in Karlsruhe indiziert. Laut Dokumentation werden deutsche Umlaute unterstützt, suchen mit Umlauten führen aber momentan nicht zu einem Ergebnis. Dies kann daran liegen, dass der Server unter Digital Unix dies nicht richtig unterstützt.

Versionen: Source Code. Kann z.B. mittels GNU C++ übersetzt werden.

Lizenzen: frei verfügbar (GNU Lizenz)

Informationen / Download: <http://www.htdig.org/>

Auszug aus einem Vergleichstest (<http://www.heise.de/ix/artikel/1999/02/089/default.shtml>):

ht://Dig

Diese Programmpaket zählt zu den HTTP-basierten Indizierern und arbeitet sich rekursiv durch einen oder mehrere WWW-Server. Zur Erhöhung der Trefferquote lassen sich bei ht://Dig [m] eigene, zusätzliche META-Tags in die HTML-Seiten integrieren, die dann entsprechend in die Datenbank aufgenommen werden (siehe 'Metainformationen'). ht://Dig liegt als Quelltext vor und ist unter den Rahmenbedingungen der GNU Public License (GPL) frei verfügbar. Die letzte Änderung auf den WWW-Seiten datiert vom Dezember1998.

Das Einrichten des Paketes verlief dank eines configure-Scripts ohne größere Überraschungen, dauerte allerdings recht lange. Nur eine manuelle Überarbeitung der config-Datei und der Makefiles im Sourcebaum waren notwendig. Bereits beim Übersetzungsprozeß macht sich der größere Leistungsumfang als bei Excite oder WebGlimpse bemerkbar: das Programm sucht nach dem PDF-Betrachter acroread und nach diversen Zusatzwerkzeugen. ht://Dig beherrscht sogar eine eigene unscharfe Suche.

Nach kurzer und minimaler Anpassung der Konfigurationsdatei ht://Dig.conf baut der Befehl rundig eine Datenbank auf. Mit dabei ist ein vorkonfiguriertes Abfrageformular, das aber nur wenige Optionen unterstützt. Die ausgezeichnete Online-Dokumentation läßt allerdings kaum Wünsche offen und zeichnet Wege für die eigenen Anpassungen, unter anderem über Templates, Umgebungsvariablen und so weiter.