FIZ CHEMIE BERLIN

Fachinformationszentrum Chemie GmbH

Jost T. Bohlen, René Deplanque, Michael Langner

The ChemGuide Access to Chemistry Information on the Web

ACS Fall Meeting, Chicago, August 27, 2001

Average Number of Hosts (Netsizer)

January 1999	44,281,500
January 2000	70,339,700
January 2001	105,959,000
June 2001	125,709,000

Growth: > 35 mill. per year

Number of Pages ???

Harris Poll #19, April 2001

- 52% of US citizens accessing health care information (i.e. 75% of all adults online) use portals or search engines
- 24% go directly to a site focused on health-related topics
- 16% go first to a general site that focuses on many topics and may have a section on health issues

Portal

<*World-Wide Web*> A web site that aims to be a "doorway" to the World-Wide Web, typically offering a search engine and/or links to useful pages, and possibly news or other services. These services are usually provided for free in the hope that users will make the site their default home page or at least visit it often. Most portals on the Internet exist to generate advertising income for their owners, others may be focused on a specific group of users and may be part of an intranet or extranet.

FOLDOC, Copyright 1993 by Denis Howe, Imperial College of Computing

FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web

- **Directories**
  - list (registered) sites
  - clustered by topic
- **Search Engines**
  - spider based
  - try to cover the entire web
- **Link Lists**
  - topic related
  - individually maintained

FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web

- **Yahoo**
  - Category 'Organic Chemistry' lists only 49 sites
- **AltaVista**
  - Search for 'Organic near Chemistry' retrieves > 370,000 pages
- **Google**
  - Search for 'Organic Chemistry' retrieves >40.000 pages

FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web

- **Directories**
  - arbitrary results
  - few hits
- **Search Engines**
  - too many (off topic) hits
  - almost impossible to evaluate
- **Link Lists**
  - time consuming to follow
  - often outdated

FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web

The 'Guide' Concept

- Pre-selected and topic-related Web sites
- Automated frequent updates
- Full-text indexing
- Sophisticated retrieval
- Combined searches in bibliographic databases and the Internet

The ChemGuide Access to Chemistry Information on the Web

The 'Guide' Concept - Procedure

1. Start with a list of relevant URL's
2. Robots collect all entry pages and all links
3. Linked pages on the same server are collected as well
4. Collected content is indexed and stored

The ChemGuide Access to Chemistry Information on the Web

The 'Guide' Concept - Procedure

5. Links to external servers are parsed and those servers identified.
6. If an external server is cited by five or more different servers it is added to a list for evaluation.
7. New URL's are individually evaluated by scientists and added to the pool if considered relevant.

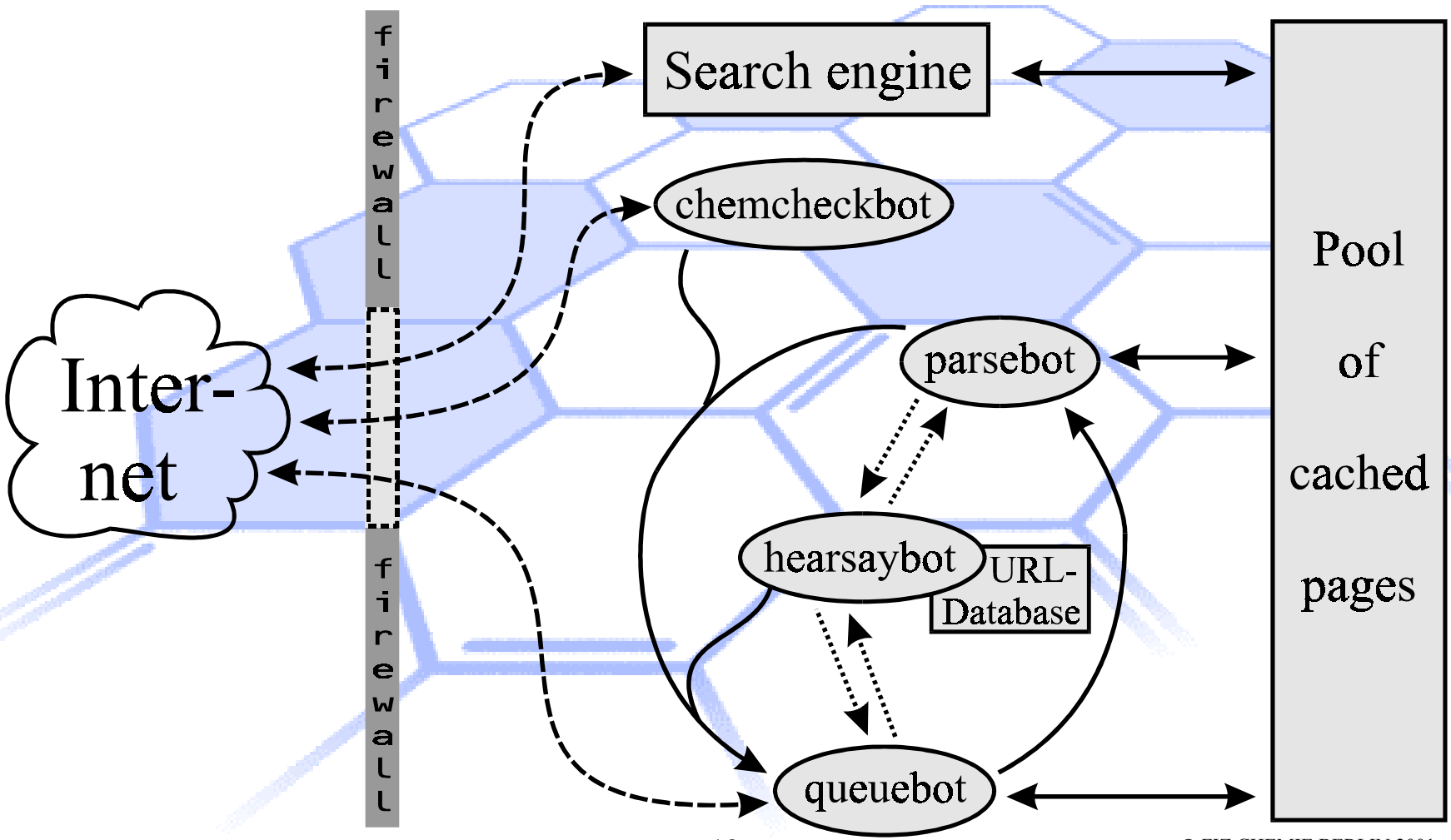
FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web

- **ChemGuide**
  - ~ 3.0 mill. chemistry URL's
- **MedPharmGuide**
  - ~ 3.0 mill. URL's featuring pharmacology, medicine, etc.
- **PublishersGuide Science and Technology**
  - ~ 1.5 mill. URL's from publishers, scientific organisations and societies

FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web



The ChemGuide Access to Chemistry Information on the Web

Retrieval

- Searches document text and/or meta tags and/or host
- Full index support for all search fields
- Versatile truncation options
- Boolean and proximity operators
- List management in expert search mode
- SDI's
- Hit terms are highlighted within retrieved documents

The ChemGuide Access to Chemistry Information on the Web

- **Truncation**
  - **? Left and right truncation**
    any number of characters
  - **# Right truncation with one or zero characters**
    vapor# = vapor or vapors
  - **! Exactly one character within or at the end of a search term**
    sulf!!yl = sulfenyl or sulfoxyl

The ChemGuide Access to Chemistry Information on the Web

- **Operators**
  - **Boolean**

    And, Or, Not
  - **Proximity**

    With, Adjacent

    (nW) liquid(w)crystal

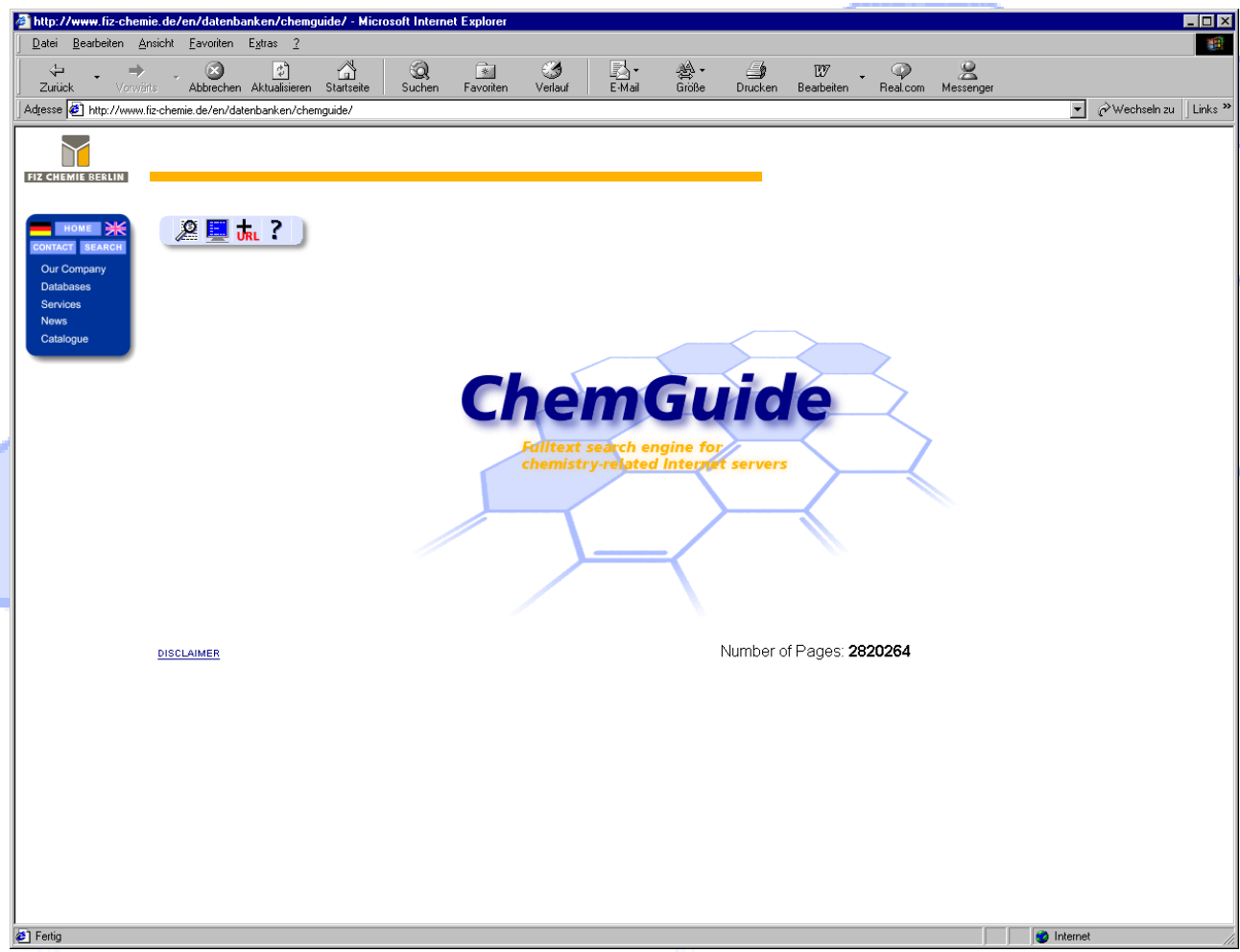
    (nA) system(a)information

    = 'system information' or 'information system'

n = maximum number of words between search terms

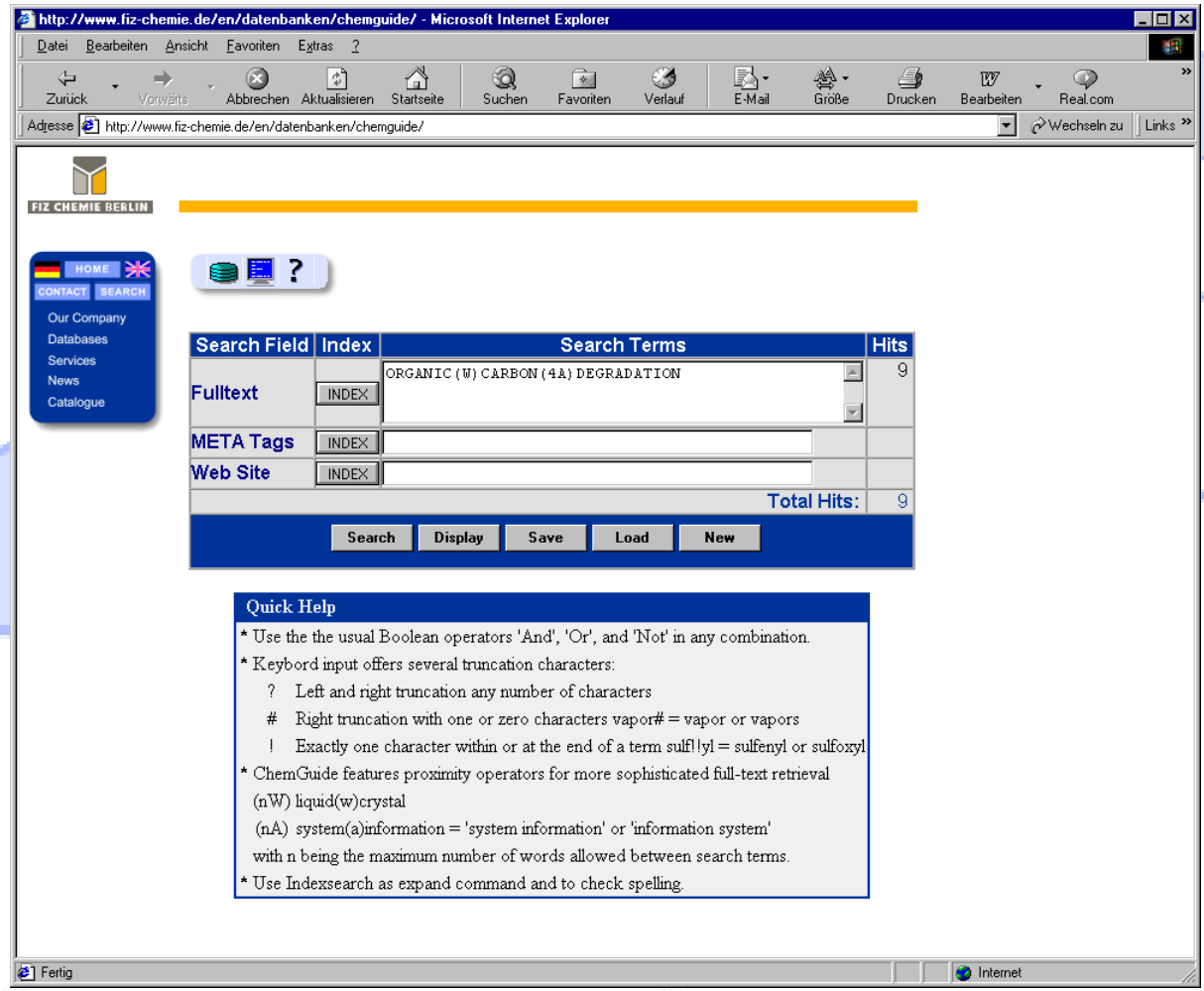
FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web



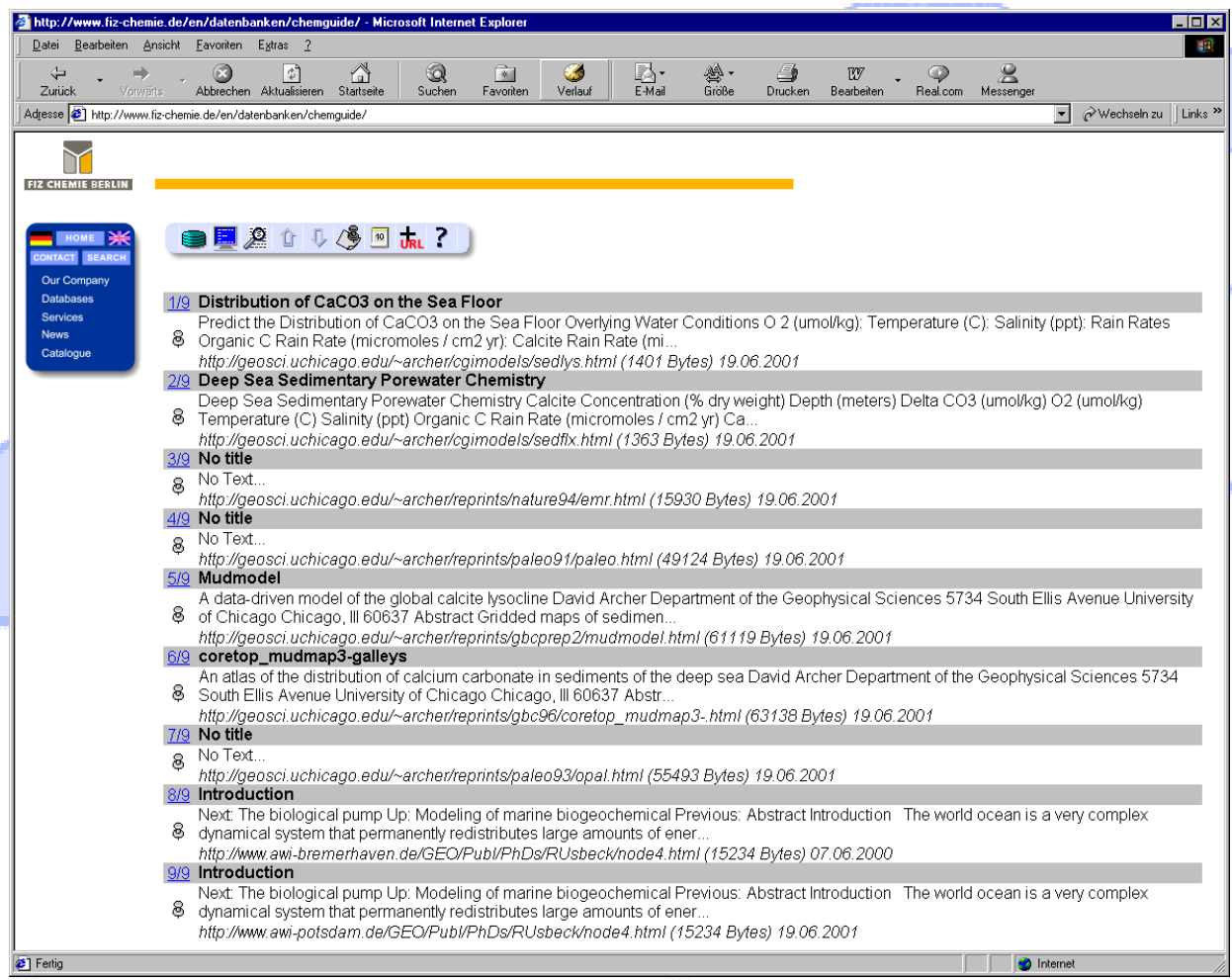
FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web



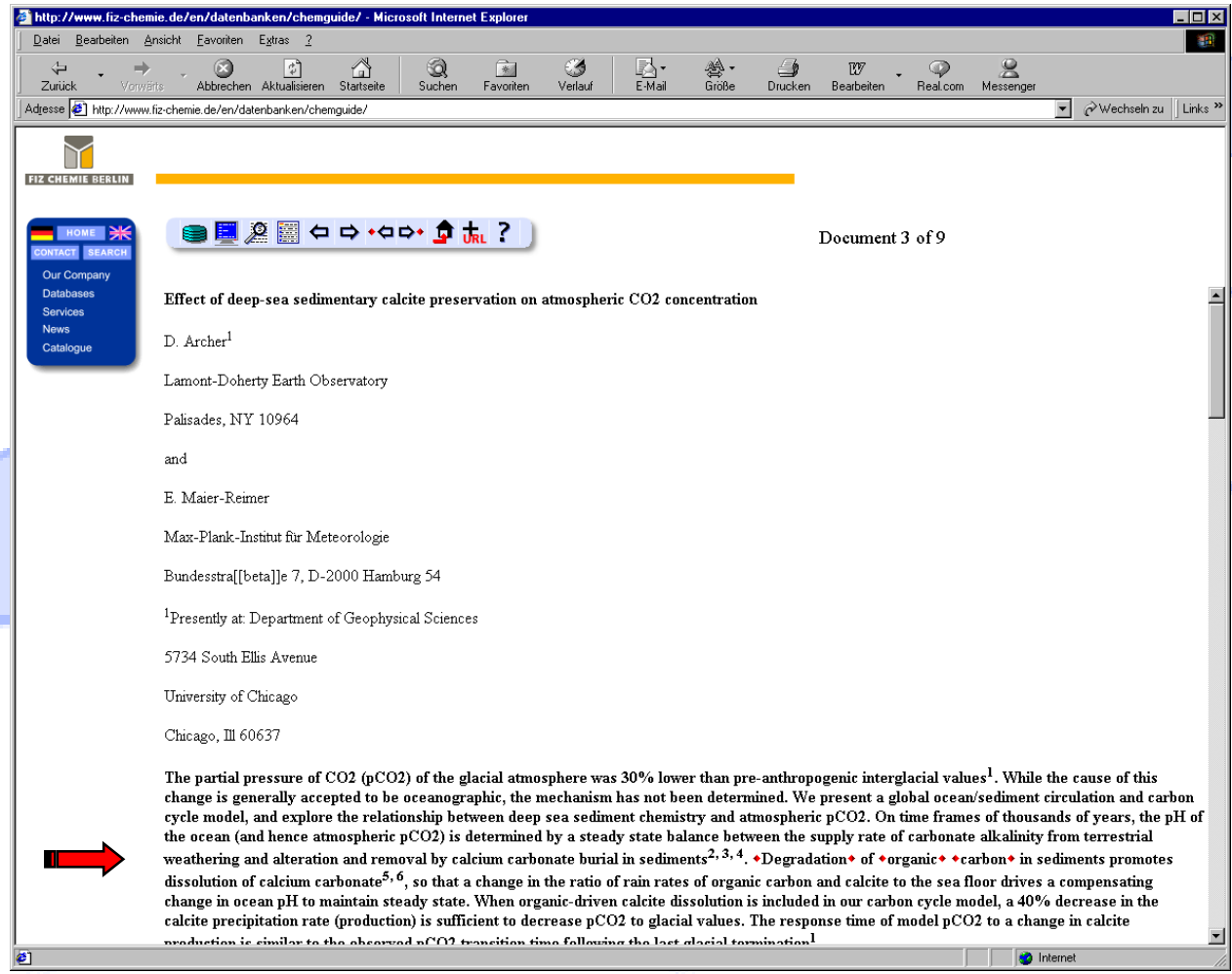
FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web

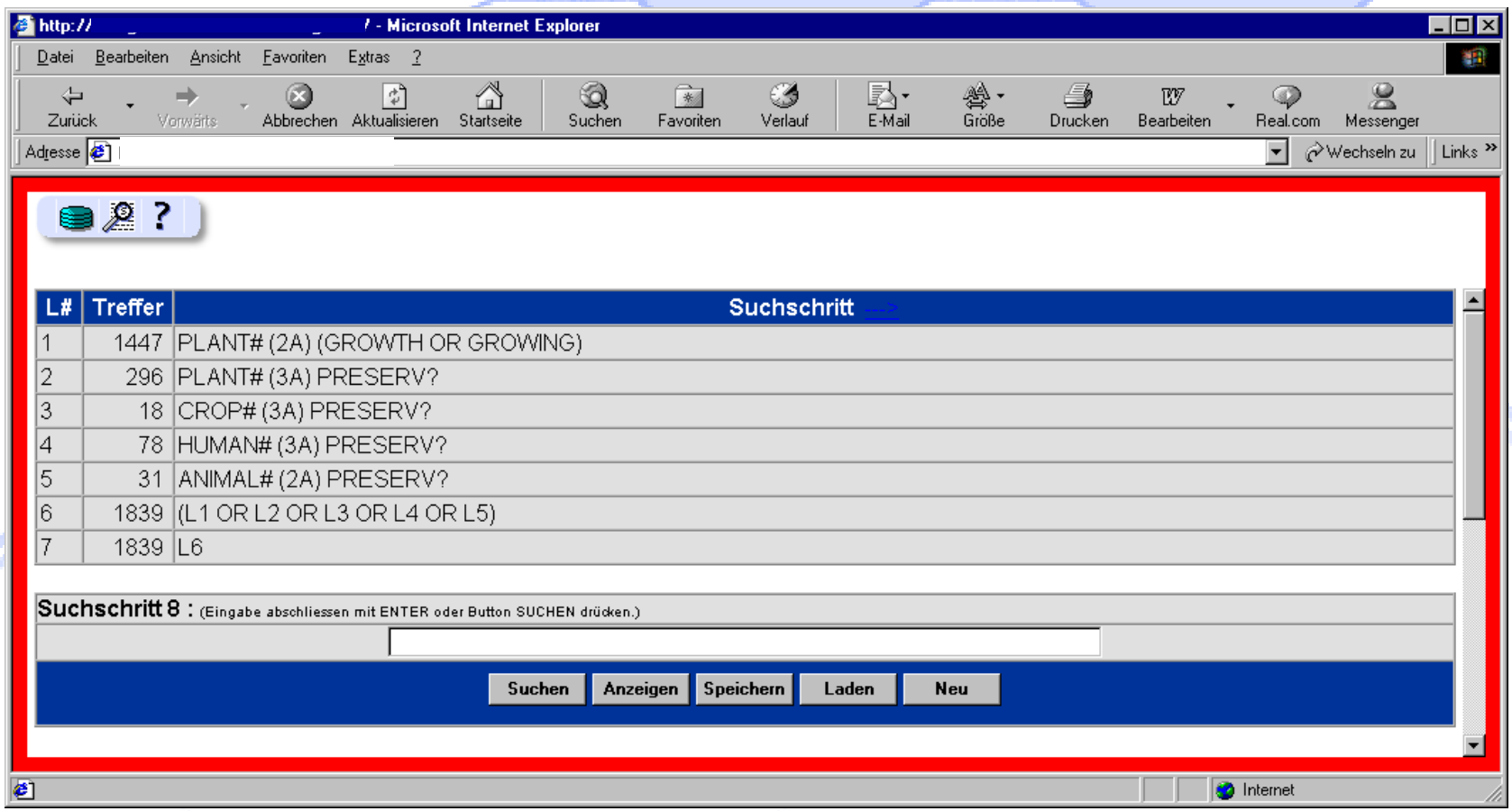


FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web



Expert Search - ChemGuide



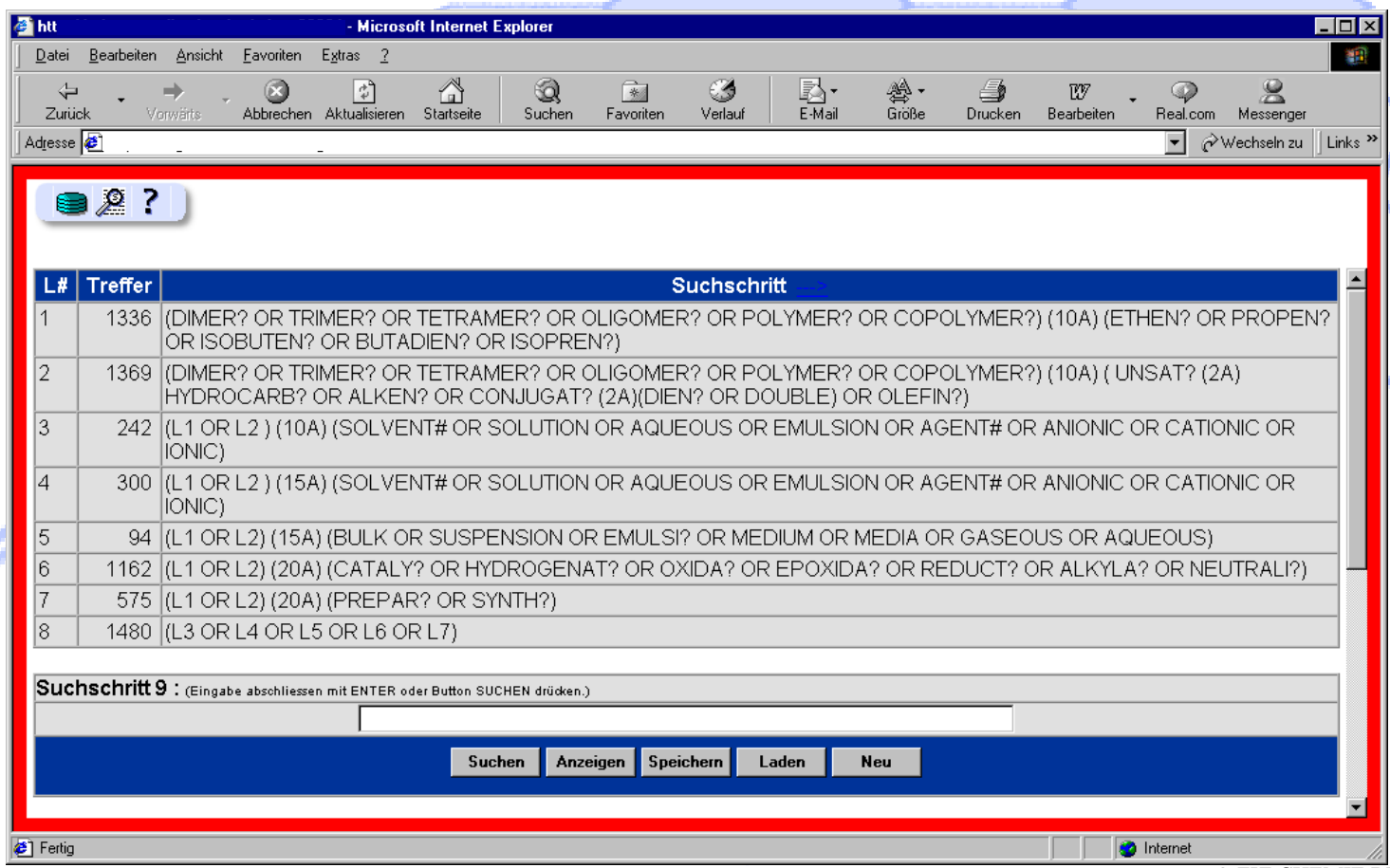
The ChemGuide Access to Chemistry Information on the Web

- **SDI (Selective Dissemination of Information)**
  - repetitive searches with identical queries
  - complex search profiles
  - run with every update
  - results editable and distributable via e-mail

FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

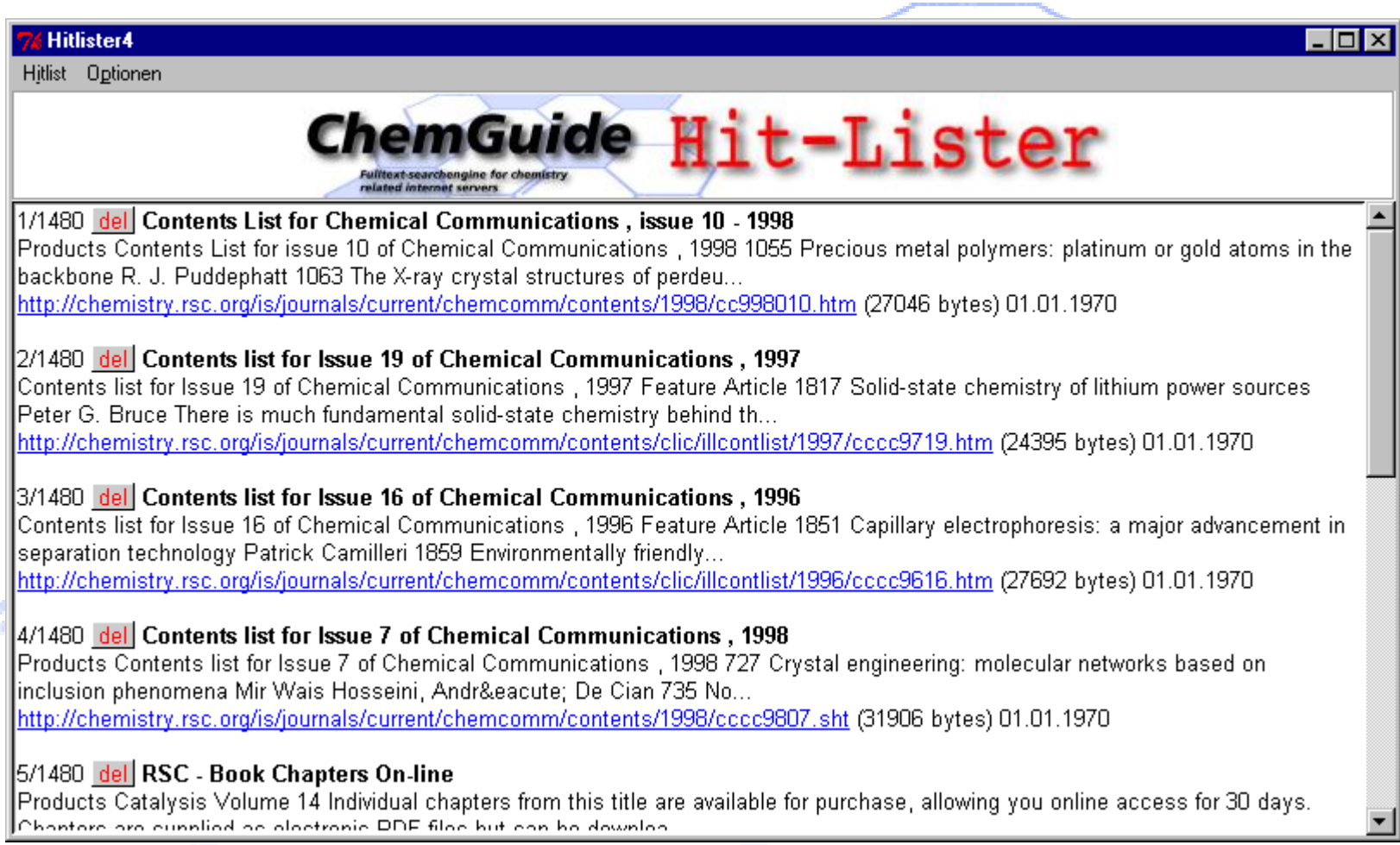
The ChemGuide Access to Chemistry Information on the Web

SDI-Example ChemGuide



FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web



The ChemGuide Access to Chemistry Information on the Web

From Guide to Portal

- Create a customized search engine for given topic
- Generate a set of complex queries that produce hits for special points of interest
- Connect queries to corresponding search buttons
- Add additional features (link lists, discussion groups, billboards, etc.)

The ChemGuide Access to Chemistry Information on the Web

New Portals

- Applicable to any topic, set of hosts, or Intranet
- Automatic frequent updates
- Minimal maintenance after initial installation
- Answer sets for areas of interest are dynamically generated by fixed queries. Thus, they always resemble current content.
- SDI's available for registered users

FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

The ChemGuide Access to Chemistry Information on the Web