ECIR'08 Industry Day

# Guided Summarization

Daniel Tunkelang
Chief Scientist, Endeca

© 2008 Endeca Technologies, Inc. All rights reserved.

# overview

- Who is Endeca?
- What is guided summarization?
- Why should you care?

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# Who is Endeca?

- Software to help people explore, analyze, and understand complex information, guiding them to unexpected insights and better decisions.

- 500+ customers
  - ABN AMRO, Boeing, Cox, Defense Intelligence Agency, Dell, Ford, Hyatt, IBM, John Deere, Library of Congress, Texas Instruments, Wal-Mart

- $108M revenue in 2007.

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# some of our customers

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECR'08

# some European customers

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECR'08

# What is guided summarization?

- Set retrieval that responds to queries with
  - an overview of the user's current context.
  - an organized set of options for incremental exploration.

- Contextual summaries of document sets optimize system’s communication with user.

- Query refinement options optimize user’s communication with system.

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# Why should you care?

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# why you should care

- Search has hit a wall.
- Guided summarization succeeds where search fails.
- Guided summarization is more than faceted search.

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

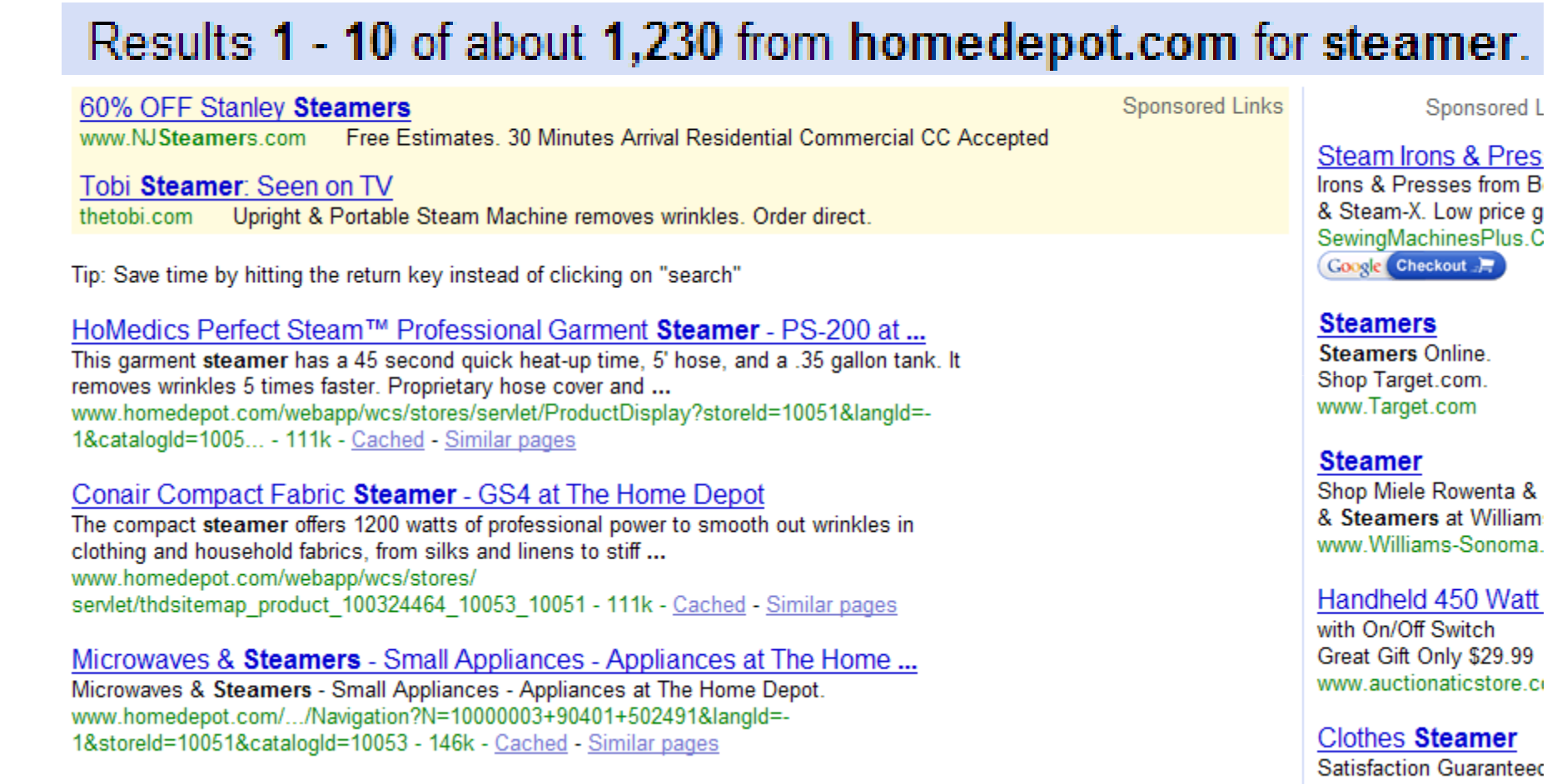
Search has hit a wall.

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# search hits a wall in ecommerce




© 2008 Endeca Technologies, Inc. All rights reserved.

# search hits a wall in KM

## Current Search: it outsourcing

Hide Summaries
Print Results

**Showing Results:** 1-25 1 | next page

**Sort by Relevancy** Sort by Date

**Report Highlight for User Survey Analysis: IT Outsourcing Becomes Mainstream in Western Europe**
9 January 2008 Pages: 3
Markets

Gianluca Tramacere

Data collected in 2007 from 356 Western European respondents can help you to better understand users' practices, opinions, drivers, inhibitors, competency levels and satisfaction with IT outsourcing.

Click to Browse:
Applications Outsourcing; Data Center Services; Desktop Services; Help Desk Services; IT Outsourcing; IT Services Marketplace; Network Services

**Report Highlight for Dataquest SWOT: EDS, IT Outsourcing Services, Worldwide**
25 April 2007 Pages: 3
Vendors

Dane S. Anderson | Kurt Potter | Gianluca Tramacere | Allie Young | Minjoo Chon

Gartner examines EDS's worldwide IT outsourcing services through an analysis of strengths, weaknesses, opportunities and threats.

Click to Browse:
IT Outsourcing; IT Services Marketplace

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# search even hits a wall on the web

Results **1-10** out of about **344,000,000** for **ir**

Tip: Save time by hitting the return key instead of clicking on "search"

**IR** (INGERSOLL RAND LTD A)

Google Finance  Yahoo Finance  MSN Money  MarketWatch  CNN Money  Reuters

**43.15** -1.13 (-2.55%) Mar 14 4:01pm ET

| | | | |
|---|---|---|---|
| Open: | 44.51 | Volume: | 5,137,478 |
| High: | 44.73 | Avg Vol: | 5,584,000 |
| Low: | 42.35 | Mkt Cap: | 11.76B |

NYSE data delayed by 20 minutes - Disclaimer

After Hours: **43.00** -0.15 (-0.35%) Mar 14 4:38pm ET

**International Rectifier** - The Power Management Leader
Manufacturer of power semiconductors (MOSFET, IBGT, Diodes and Thyristors).
Stock quote for IRF
www.**ir**f.com/ - 27k - Cached - Similar pages

**Infrared** spectroscopy - Wikipedia, the free encyclopedia
**Infrared** spectroscopy (**IR** spectroscopy) is the subset of spectroscopy that deals with the **infrared** region of the electromagnetic spectrum. ...
en.wikipedia.org/wiki/**Infrared**_spectroscopy - 54k - Cached - Similar pages

© 2008 Endeca Technologies, Inc. All rights reserved.

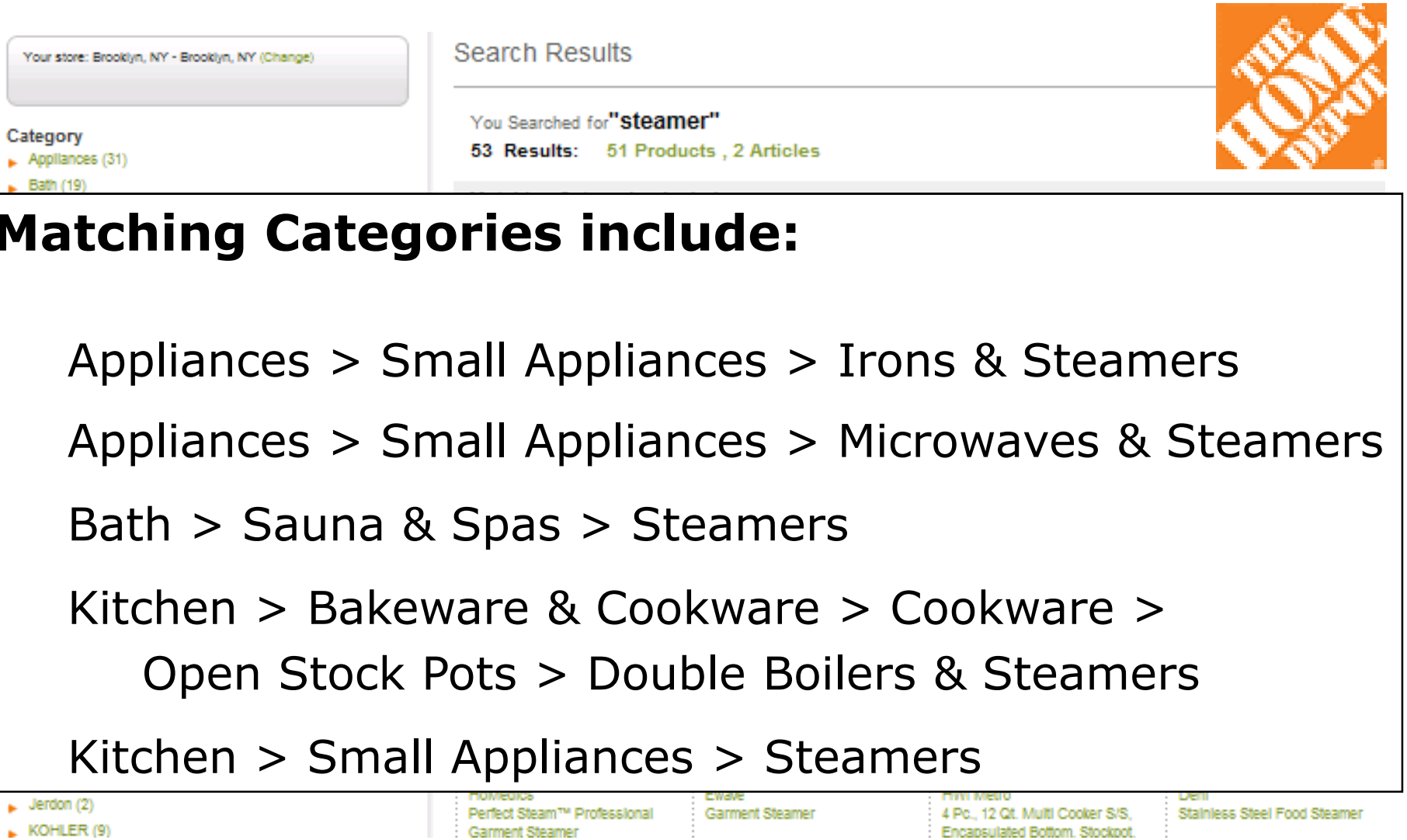
Guided summarization succeeds where search fails.

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# guided summarization for ecommerce

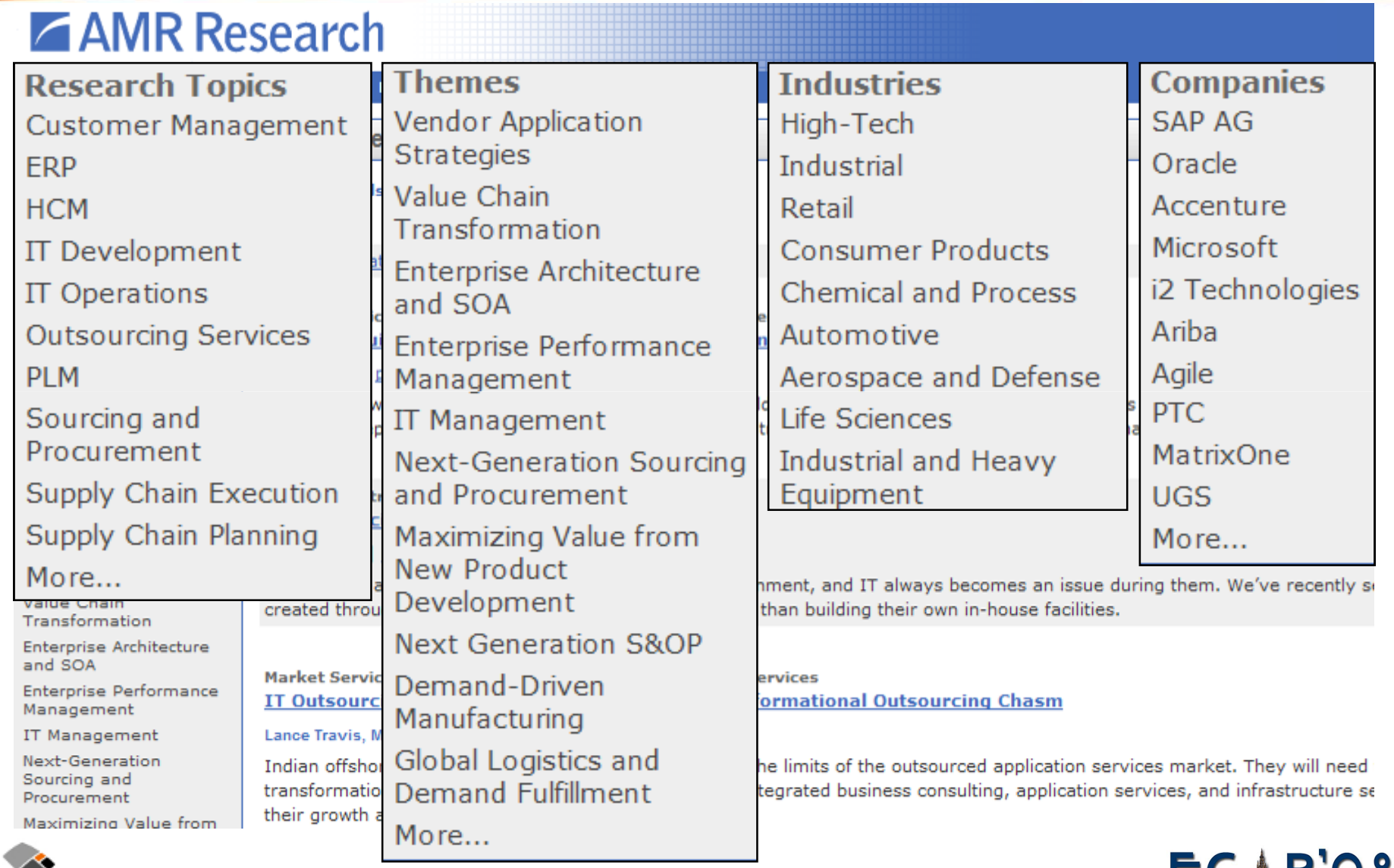



ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# guided summarization for KM




ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

Guided summarization is more than faceted search.

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# facets 101

## Flamenco UC Berkeley Architecture Slides

A snapshot of Images from the UC Berkeley Architecture Visual Resources Library

Save Search | History and Settings | Return to

"gehry" appears in these category names:

- **Source** > Arnell, Peter and Bickford, Ted, eds - Frank Gehry, buildings and projects
- **People** > Gehry, Frank

search

all items / in current results

These terms define your current search. Click the × to remove a term.

keyword "gehry" ×

Refine your search within these categories:

**PEOPLE** (group results)

architect (94)

**PERIODS** (group results)

20th Century (93) | Modern (1)

**LOCATIONS** (group results)

North America (35) | Western Europe (58)

**STRUCTURE TYPES** (group results)

architectural elements (6)
buildings (by function) (84)
buildings (by massing or shape) (1)
details (6)
multi-building complexes (14)
parts of buildings (9)
transportation elements (3)

**Items 1 to 40 of 94 results**

1 41 81

Gehry House

Gehry House

Gehry House

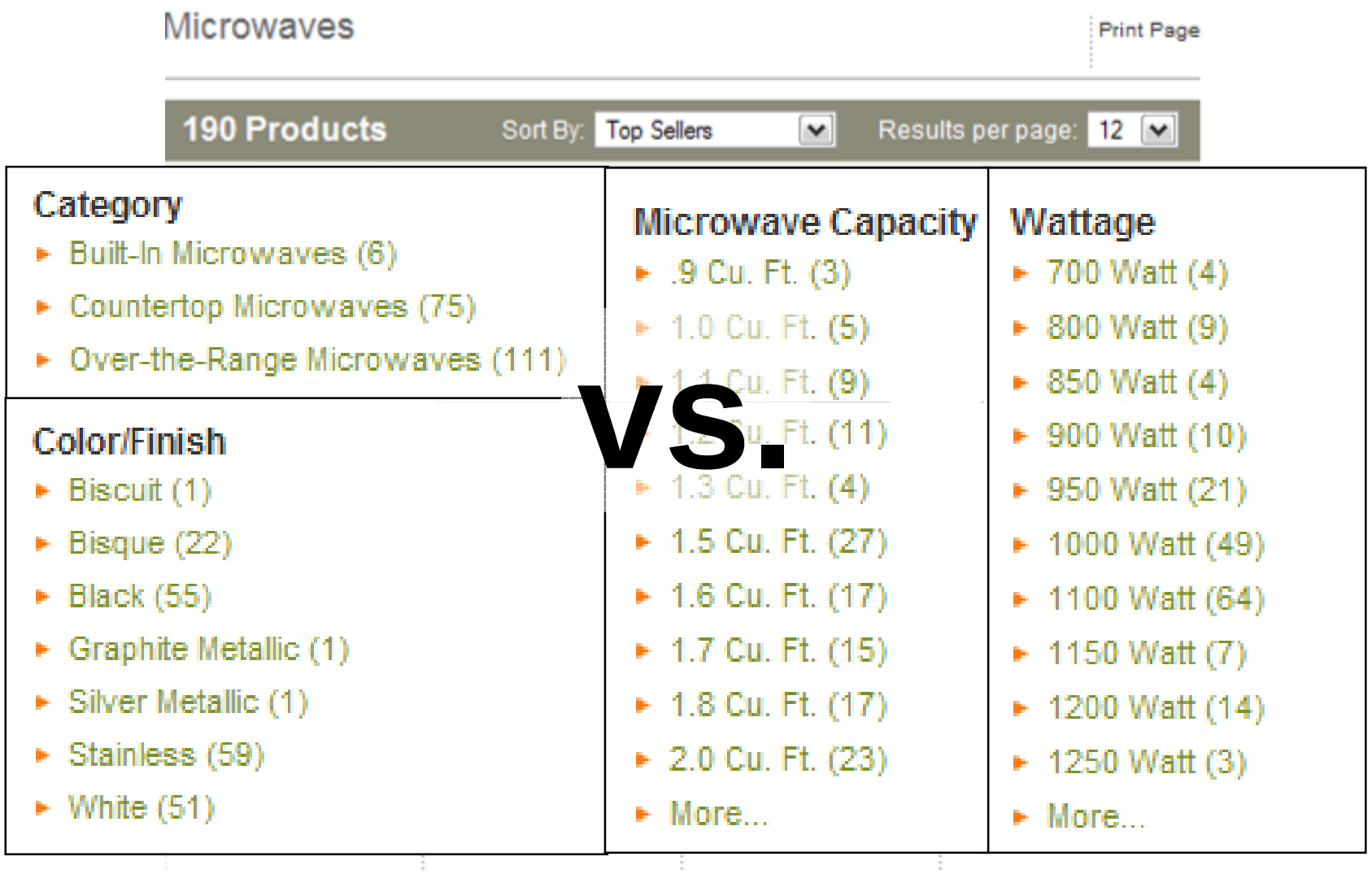
Frank Gehry Offices

Frank Gehry Offices

Frank Gehry Offices

© 2008 Endeca Technologies, Inc. All rights reserved.

# showing the right facets: microwaves

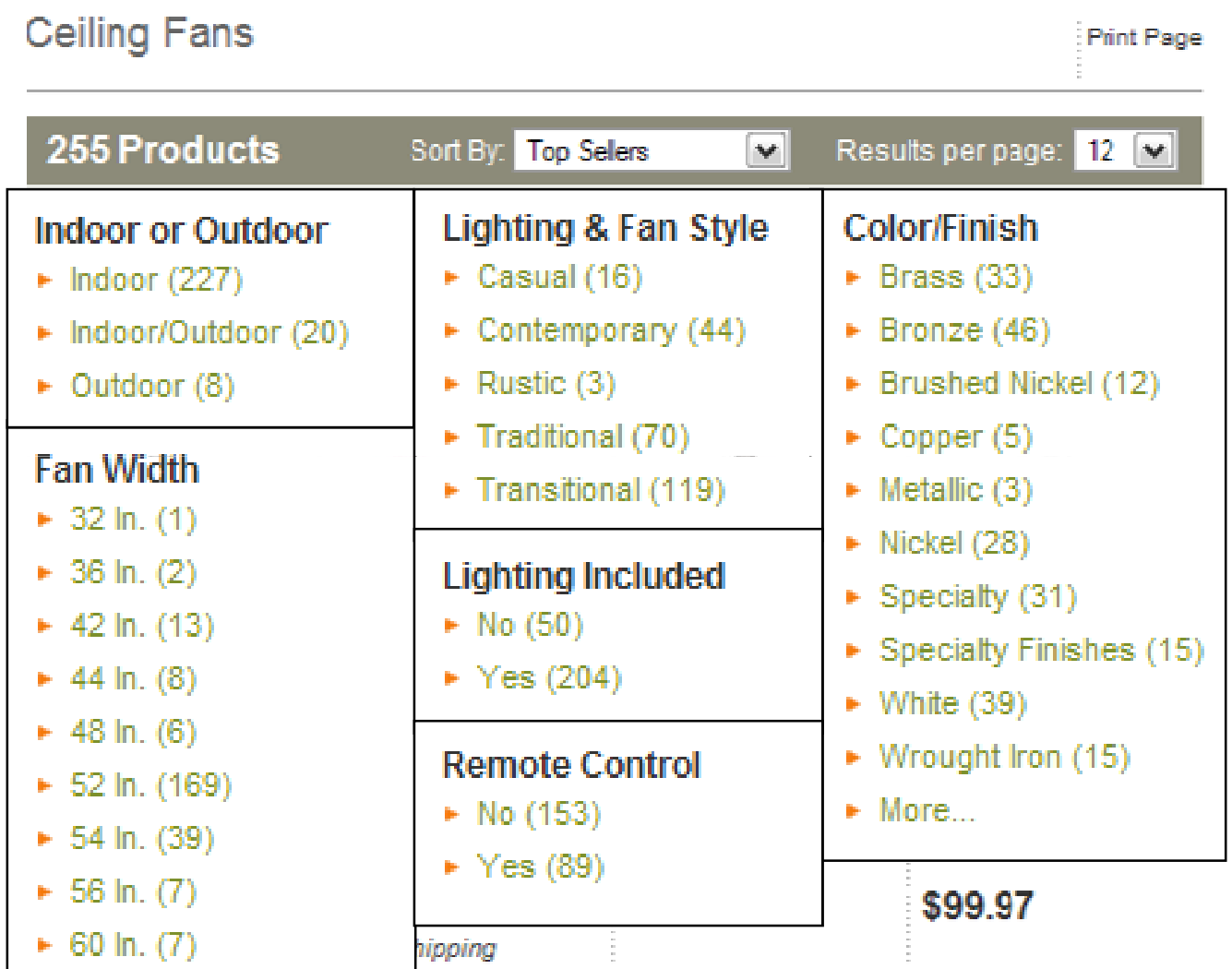



ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# showing the right facets: ceiling fans

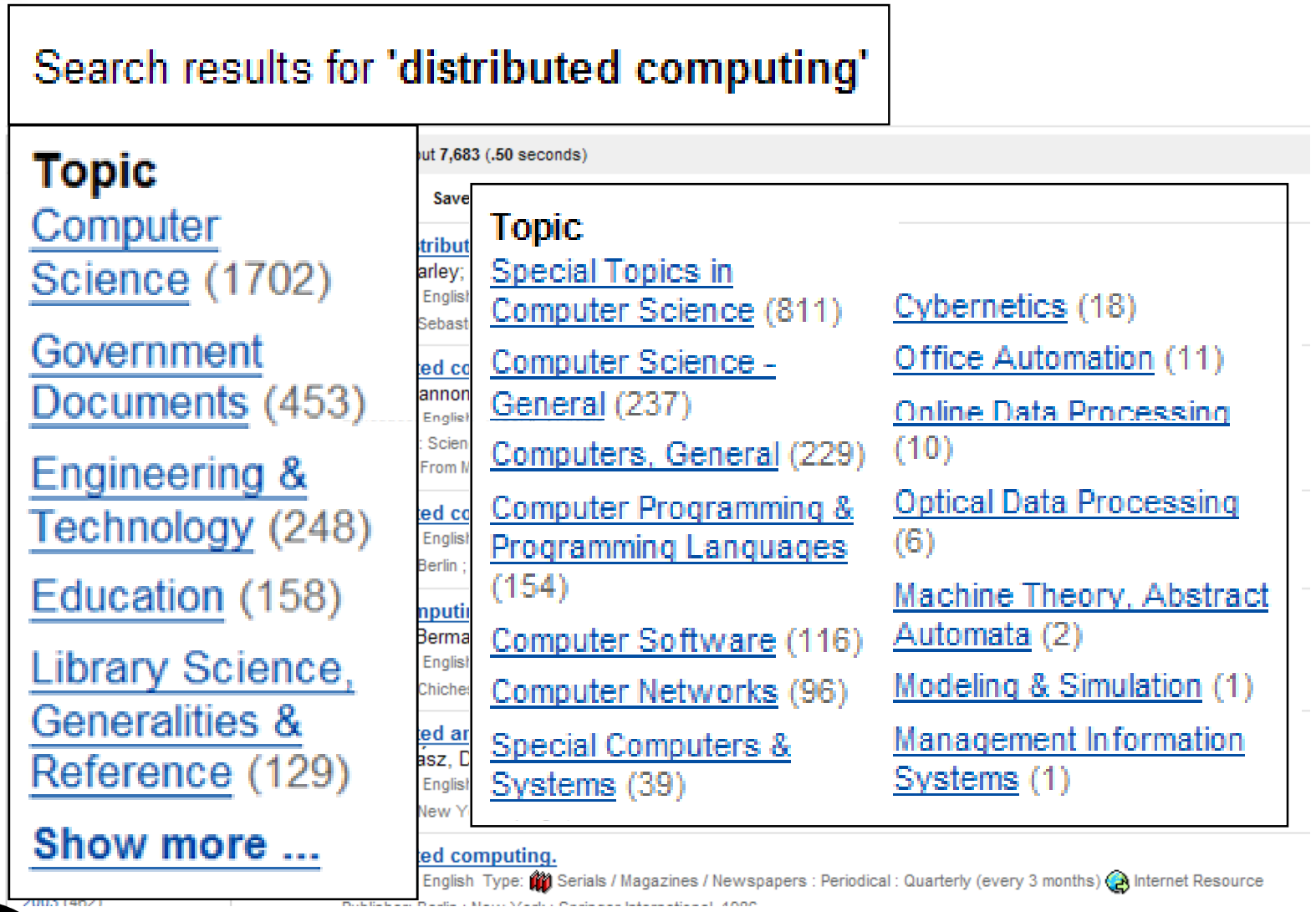



ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# traditional topic taxonomy




© 2008 Endeca Technologies, Inc. All rights reserved.

# dynamic topic facet

## Subject

- Artificial intelligence (227)
- Automatic theorem proving (9)
- Client/server computing (185)
- Computer algorithms (110)
- Computer architecture (162)
- Computer networks (552)
- Computer programs (139)
- Computer security (151)
- Computer software (253)
- Computers (124)
- Database management (277)
- Distributed processing (937)
- Electronic data processing (1002)
- Electronic digital computers (148)
- Fault-tolerant computing (365)
- High performance computing (244)
- History (11)
- Information technology (145)
- Java (77)
- Law and legislation (70)
- Logic, Symbolic and mathematical (16)
- Mathematics (70)
- Mobile communication systems (54)
- Operating systems (87)
- Parallel processing (619)
- Research (83)
- Software engineering (197)
- Supercomputers (139)
- Web databases (54)
- Wireless communication systems (97)

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# cutting through facets to show the big picture

Might we suggest that you refine by...

V Toys: Chests & Storage

Disney Princess Toy Organizer

Disney Princess Toy Box

Little Tikes Mission Design Toy Box

V Category: Computers

EDGE 128 MB DiskGO! USB Flash Drive

Comstar External DVD+R/RW Drive With USB 2.0 CSDVRW02E

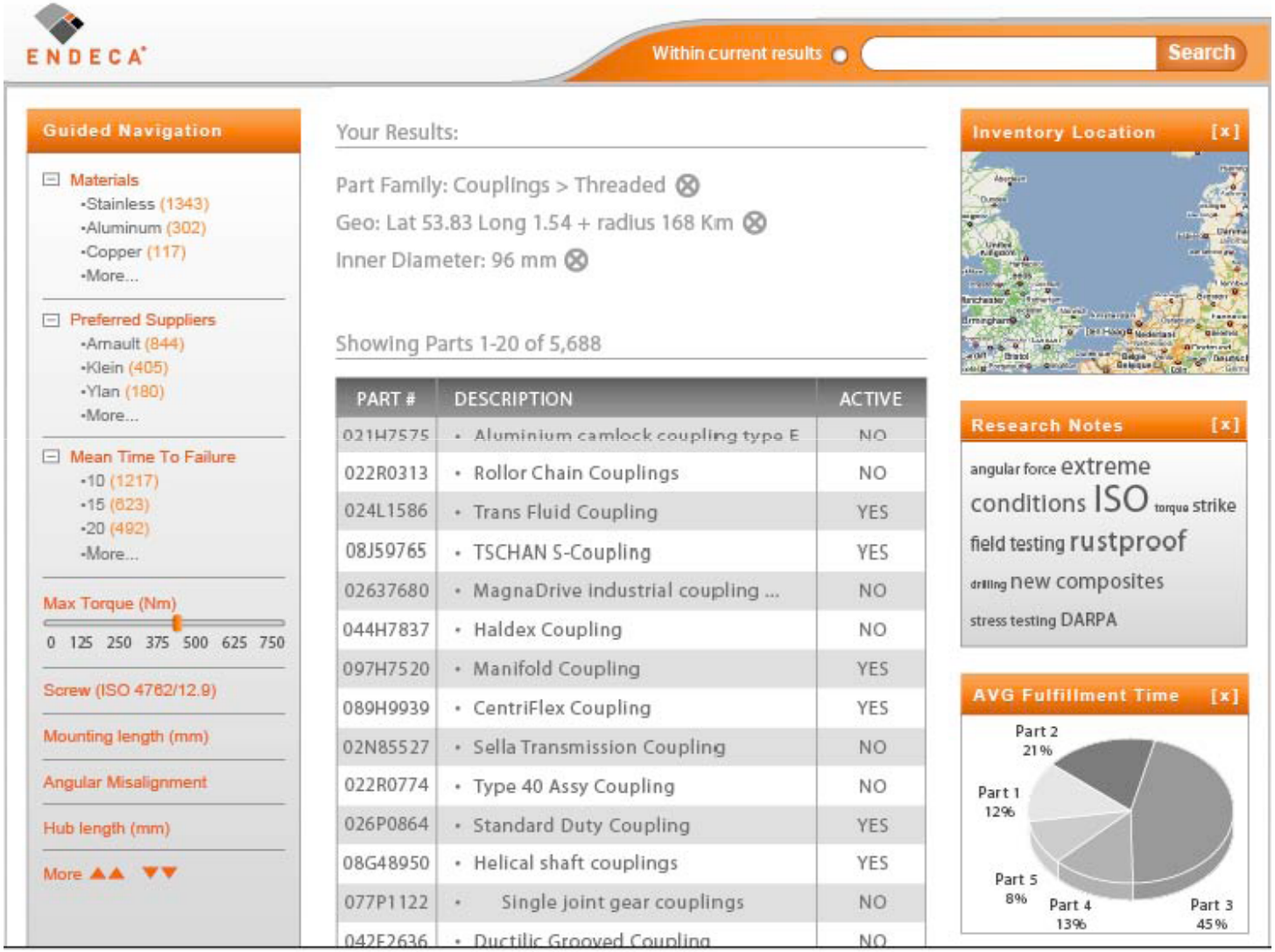
SanDisk 256 MB Memory Stick Pro

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# summarization: more than search and browse




ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# guided summarization – a summary

Guided summarization enables a dialog between the user and the data, enabling exploration and discovery.

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08

# thank you

Questions?

ENDECA

© 2008 Endeca Technologies, Inc. All rights reserved.

ECIR'08