Faceted Search

Synthesis Lectures on Information Concepts, Retrieval, and Services

Editor
Gary Marchionini, *University of North Carolina, Chapel Hill*

Faceted Search
Daniel Tunkelang
2009

Introduction to Webometrics: Quantitative Web Research for the Social Sciences
Michael Thelwall
2009

Automated Metadata in Multimedia Information Systems: Creation, Refinement, Use in Surrogates, and Evaluation
Michael G. Christel
2009

Exploratory Search: Towards New Search Horizons
Ryen W. White and Resa A. Roth
2009

Reference in a Digital Age
R. David Lankes
2009

© Springer Nature Switzerland AG 2022

Reprint of original edition © Morgan & Claypool 2009

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Faceted Search
Daniel Tunkelang

ISBN: 978-3-031-01134-4 paperback

ISBN: 978-3-031-02262-3 ebook

DOI: 10.1007/978-3-031-02262-3

A Publication in the Springer series

SYNTHESIS LECTURES ON INFORMATION CONCEPTS, RETRIEVAL, AND SERVICES

Lecture #5

Series Editor: Gary Marchionini, University of North Carolina

Series ISSN

ISSN 1947-945X print

ISSN 1947-9468 electronic

Faceted Search

Daniel Tunkelang
Endeca

SYNTHESIS LECTURES ON INFORMATION CONCEPTS, RETRIEVAL, AND SERVICES # 5

ABSTRACT

We live in an information age that requires us, more than ever, to represent, access, and use information. Over the last several decades, we have developed a modern science and technology for information retrieval, relentlessly pursuing the vision of a "memex" that Vannevar Bush proposed in his seminal article, "As We May Think."

Faceted search plays a key role in this program. Faceted search addresses weaknesses of conventional search approaches and has emerged as a foundation for interactive information retrieval. User studies demonstrate that faceted search provides more effective information-seeking support to users than best-first search. Indeed, faceted search has become increasingly prevalent in online information access systems, particularly for e-commerce and site search.

In this lecture, we explore the history, theory, and practice of faceted search. Although we cannot hope to be exhaustive, our aim is to provide sufficient depth and breadth to offer a useful resource to both researchers and practitioners. Because faceted search is an area of interest to computer scientists, information scientists, interface designers, and usability researchers, we do not assume that the reader is a specialist in any of these fields. Rather, we offer a self-contained treatment of the topic, with an extensive bibliography for those who would like to pursue particular aspects in more depth.

KEYWORDS

faceted search, exploratory search, information seeking, human–computer information retrieval

Preface

We live in an information age, a world where, more than ever, we need to understand how to represent, access, and use information. Over the last several decades, we have witnessed the development of a modern science and technology for information access. We've come a long way from the vision of a "memex" that Vannevar Bush proposed in his seminal article, "As We May Think" [1]: a mechanical device that would allow someone to access a large, self-contained research library.

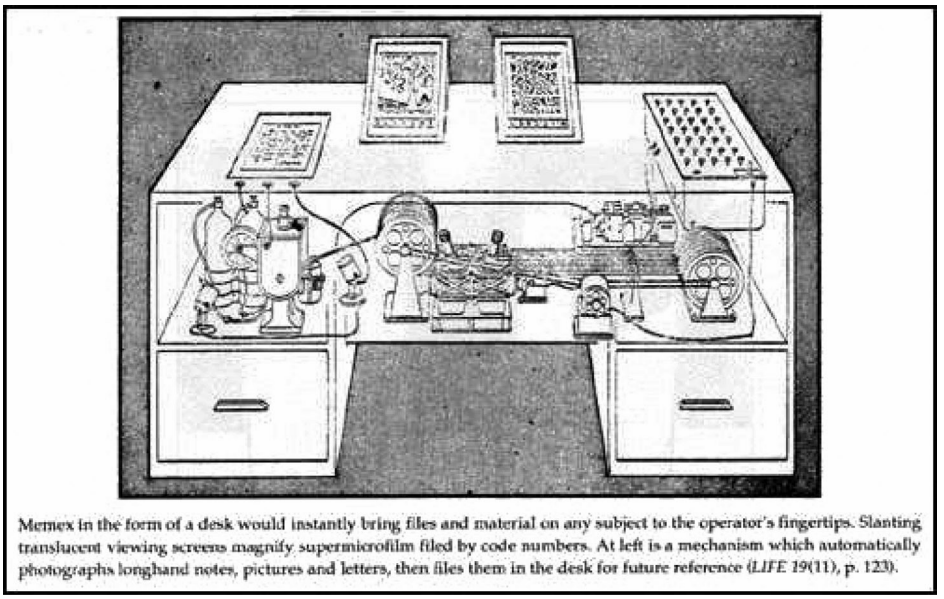
Many people may feel that we have already achieved that end goal, albeit by different means. After all, we can now enter the name of a person or company into Google or some other web search engine and, in most cases, be instantly directed to the associated web page. Thanks to tools like the collectively edited Wikipedia, we may achieve similar success with more general queries, at least for topics of broad enough interest to have inspired Wikipedia entries.

Looking beyond these use cases, however, we see that we are only in the early days of implementing the memex vision (Figure 1). Modern search engines adequately address the problem of what library scientists have historically called known-item search: we know what we are looking for and are certain it exists in the collection we are searching [2].

In contrast, we have not developed comparably mature tools for exploratory search—that is, information seeking where users do not have a known target document and may not even have a well-established information need [4]. Only in the last few years have we seen an emerging program of human–computer information retrieval (HCIR) that brings interactive techniques—many inspired by pre-Internet research in library science—to bear on more sophisticated information-seeking tasks [5].

Facets, a way of classifying information, play a key role in this program. Faceted classification addresses the weakness of earlier knowledge representations—namely, the rigidity of taxonomical schemes and the chaos of unstructured indexes. Developed by library scientists, faceted classification offers an approach to knowledge representation that is both faithful to its richness and practical for real-world use.

Faceted classification, however, only addresses the problem of representing information. We still need a means to access and use that information. That means is faceted search.



Memex in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference (*LIFE* 19(11), p. 123).

FIGURE P.1: Vannevar Bush's theoretical memex machine [3].

Researchers such as Marti Hearst have led the way with user studies that demonstrate how faceted search provides more effective information-seeking support to users than conventional best-first search—even though users are more familiar with the latter [6].

Those dubious of the value of faceted search interfaces raise the specter of what Joshua Porter and his colleagues at User Interface Engineering call the "three-click rule,"—that is, the web design rule of thumb that no piece of content should take more than three clicks to access. Luckily, that bit of folk wisdom does not hold up to empirical study [7]. When tested in a user study, it was found that there is no correlation between the number of times users clicked and their success in finding the content they sought, and that the number of clicks is not what is important to users, but only whether or not they are successful at finding what they are seeking (Figure 2).

In the lecture that follows, we will explore the history, theory, and practice of faceted search. Although faceted search has become increasingly prevalent in online information access systems, this text is, to our knowledge, the first comprehensive treatment of the subject. Although we cannot hope to be exhaustive, our aim is to provide sufficient depth and breadth to offer a useful resource to both researchers and practitioners.

Because faceted search is an area of interest to computer scientists, information scientists, interface designers, and usability researchers, we do not assume that the reader is a specialist in any of

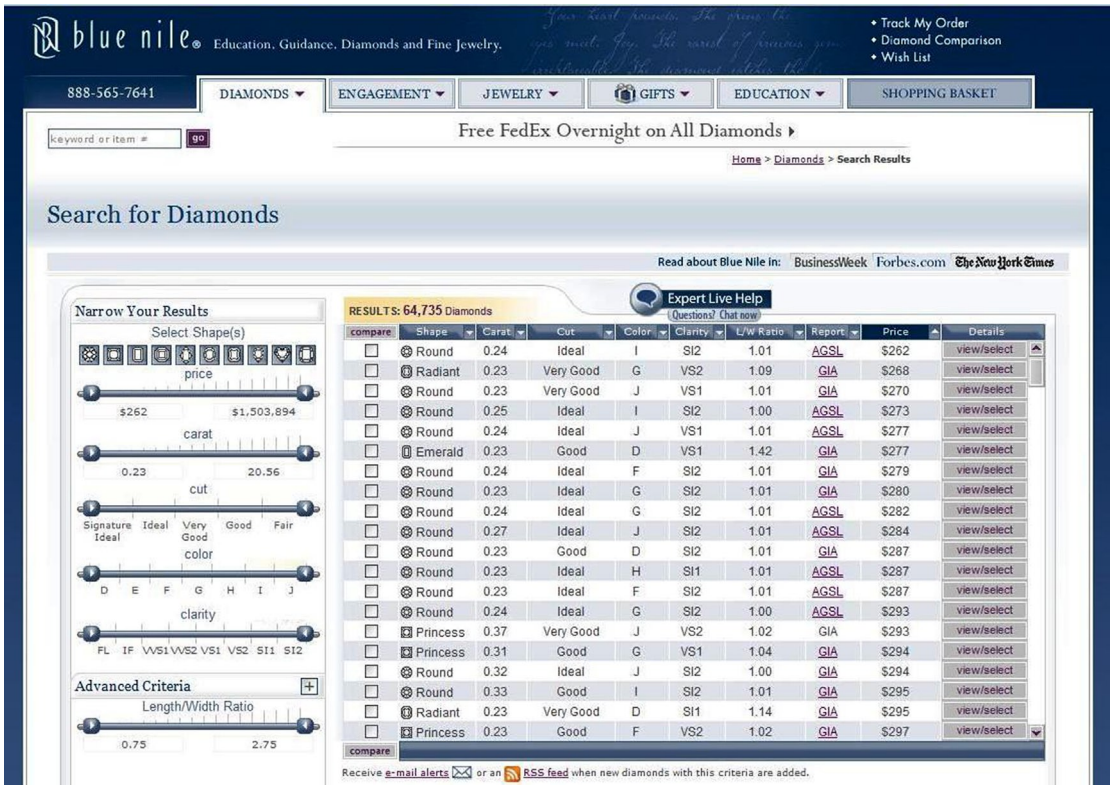


FIGURE P.2: Results for "diamonds" at www.bluenile.com.

these fields. Rather, we offer a self-contained treatment of the topic, with an extensive bibliography for those who would like to pursue particular aspects in more depth.

The book consists of three parts. The first part presents the key concepts leading the reader on a historical path from Aristotle's classical ideas of knowledge representation to a modern-day definition of faceted search. The second part describes key work on faceted search in academia and industry. The third part addresses some of the practical challenges that confront the developers of faceted search applications.

Each chapter ends with take-aways that summarize the chapter's key points. Impatient readers may skip to these, but I hope you will find the journey as valuable as the destination.

Acknowledgments

I thank Gary Marchionini, Diane Cerra, and all of the kind people at Morgan Claypool for inviting me to write this lecture. I am honored to participate in the Synthesis Lectures on Information Concepts, Retrieval, and Services, which represent a fantastic collection of work by the leading researchers in information science and retrieval.

I thank Candy Schwartz, Co-Editor-in-Chief of *Library & Information Science Research,* for reviewing an early draft of this lecture. I am also indebted to a veritable army of volunteer reviewers: Omar Alonso, Pete Bell, Amitava Biswas, Blade Kotelly, Sol Lederman, Milan Merhar, Jennifer Novosad, XiaoGuang Qi, Brett Randall, Dusan Rnic, and Joshua Young.

I thank my colleagues and management at Endeca for allowing me to take time away from my day job to try my hand as a writer. I also thank Endeca's customers, whose applications made many of the figures in this book possible.

Finally, I am grateful to my wife Kristin and my daughter Lily for supporting me and putting up with me in these manic months.

Contents

Preface ..... **vii**

Acknowledgments ..... **xi**

I. Key Concepts **1**

1. Introduction: What Are Facets? 3
   - 1.1 Classification: Aristotle's Tree 3
   - 1.2 Facets: Ranganathan's Colons 6
   - 1.3 Ontologies 8
   - 1.4 Take-Aways 9
2. Information Retrieval 11
   - 2.1 Relevance 11
   - 2.2 Set Retrieval 12
   - 2.3 Precision vs. Recall 13
   - 2.4 Ranked Retrieval 15
   - 2.5 Directory Navigation 18
   - 2.6 Take-Aways 19
3. Faceted Information Retrieval 21
   - 3.1 Parametric Search 21
   - 3.2 Faceted Navigation 23
   - 3.3 Faceted Search 24
   - 3.4 Take-Aways 26

II. Research and Practice **27**

4. Academic Research 29
   - 4.1 Dynamic Queries and Query Previews 29
   - 4.2 View-Based Search 31
   - 4.3 The Flamenco Project 33
   - 4.4 Relation Browser 35
   - 4.5 mSpace 36

4.6 Parallax 36
4.7 Take-Aways 37
5. Commercial Applications 39
5.1 Endeca 39
5.2 eBay 39
5.3 Amazon 41
5.4 Open Source 42
5.5 Take-Aways 43

III. Practical Concerns 45
6. Back-End Concerns 47
6.1 Scale 47
6.2 Efficiency 48
6.3 Information Overload 50
6.4 Availability of Metadata 52
6.5 The Vocabulary Problem 53
6.6 Multiple Entity Types 54
6.7 Take-Aways 55
7. Front-End Concerns 57
7.1 Where and When to Present Facets 57
7.2 Organizing Facets and Facet Values 60
7.3 The Search Box 62
7.4 Multiple Selection from A Facet 65
7.5 Design Patterns 66
7.6 Take-Aways 68
8. Conclusion 69

Glossary 71

References 73

Author Biography 79