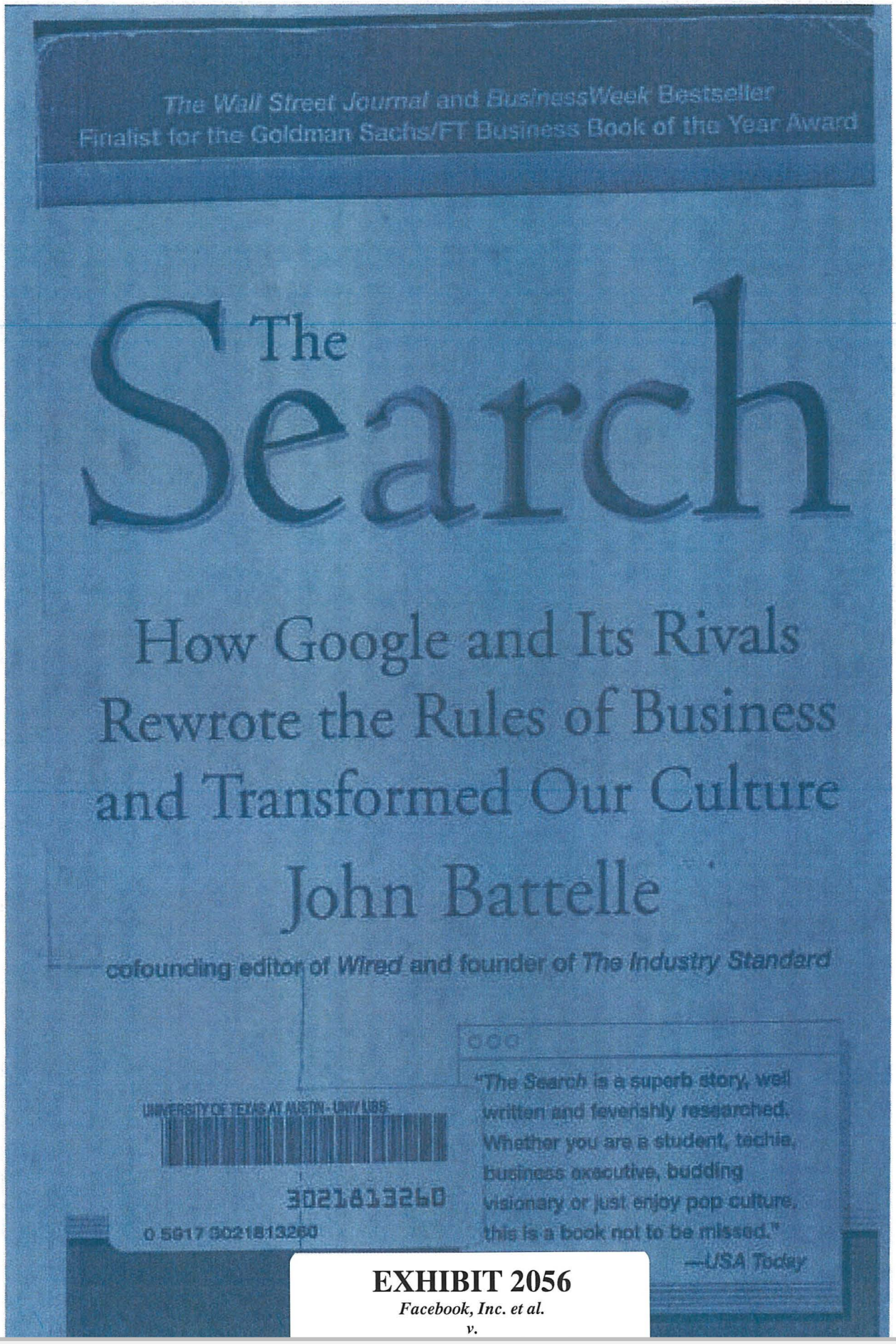

EXHIBIT 2056
Facebook, Inc. et al.
v.

For Michelle

PORTFOLIO
Published by the Penguin Group
Penguin Group (USA) Inc., 375 Hudson Street, New York, New York 10014, U.S.A. • Penguin Group (Canada), 90 Eglinton Avenue East, Suite 700, Toronto, Ontario, Canada M4P 2Y3 (a division of Pearson Penguin Canada Inc.) • Penguin Books Ltd, 80 Strand, London WC2R 0RL, England • Penguin Ireland, 25 St. Stephen's Green, Dublin 2, Ireland (a division of Penguin Books Ltd) • Penguin Books Australia Ltd, 250 Camberwell Road, Camberwell, Victoria 3124, Australia (a division of Pearson Australia Group Pty Ltd) • Penguin Books India Pvt Ltd, 11 Community Centre, Panchsheel Park, New Delhi – 110 017, India • Penguin Group (NZ), cnr Airborne and Rosedale Roads, Albany, Auckland 1310, New Zealand (a division of Pearson New Zealand Ltd) • Penguin Books (South Africa) (Pty) Ltd, 24 Sturdee Avenue, Rosebank, Johannesburg 2196, South Africa

Penguin Books Ltd, Registered Offices:
80 Strand, London WC2R 0RL, England

First published in the United States of America by Portfolio,
a member of Penguin Group (USA) Inc. 2005
This edition published 2006

10 9 8 7 6 5 4 3 2 1

THE LIBRARY OF CONGRESS HAS CATALOGED THE HARDCOVER EDITION AS FOLLOWS:
Battelle, John, 1965–
The Search : how Google and its rivals rewrote the rules of business and transformed our culture / John Battelle
p. cm.
Includes index.
Contents: The database of intentions—Who, what, where, why, when, and how (much)—Search before Google—Google is born—A billion dollars, one nickel at a time—Google 2000–2004: sees to $3 billion in five years—The search economy—Search, privacy, government, and evil—Google goes public—Google today, Google tomorrow—Perfect search.
ISBN 1-59184-088-0 (hc.)
ISBN 1-59184-141-0 (pbk.)
1. Google (Firm) 2. Internet industry—United States. 3. Web search engines.
4. Google. 5. Internet searching. 6. Information society—United States. I. Title: Google and its rivals rewrote the rules of business and transformed our culture. II. Title.
HD9696.8.U64G633 2005
338.7'610250410973—dc22 2005047538

Printed in the United States of America
Set in Adobe Garamond with Cantoll Regular • Designed by Daniel Lagin

Chapter 2

Who, What, Where, Why, When, and How (Much)

Judge of a man by his questions, rather than by his answers.
—Voltaire

Before we take a long journey around the contours and implications of search, it makes sense to get our bearings. Back when I was a cub reporter, I was taught to answer five questions about any topic before writing about it: who, what, where, why, and when. If you crammed answers to all those questions into your lead paragraph, then you'd essentially done your job.

But to those five questions I quickly learned to add a sixth—how?—and a corollary: who's making the money, and how much? We'll get to the money question last, but first, let's address the how.

How

So how does a search engine work? There's a very, very long answer to this question, but I'll stick to a shorter one. In essence, a search engine connects words you enter (queries) to a database it has created of Web pages (an index). It then produces a list of URLs (and summaries of

experimental approaches to search that are not driven by this paradigm, for the most part, every major search engine is driven by this text-based approach.

A search engine consists of three major pieces—the crawl, the index, and the runtime system or query processor, which is the interface and related software that connects a user's queries to the index. The runtime system also manages the all-important questions of relevance and ranking. All three pieces are integral to the quality and speed of the engine, and there are literally hundreds of factors in each that affect the overall search experience delivered. But the basics are pretty much the same for all the engines. As Tim Bray, a search pioneer now at Sun Microsystems, puts it in his excellent series "On Search," "The fact of the matter is that there really hasn't been much progress in the basic science of how to search since the seventies."

The search all starts with you: your query, your intent—the desire to get an answer, find a site, or learn something new. Intent drives search—a maxim I'll be repeating time and again throughout this book. We'll get into the query a bit more in the "What" section below, but on average we enter one or two short words into a query box each time we search, and we click on an average of two or so results among the millions an engine often lists. In addition, the average Web searcher conducts about one search a day. Of course, that's an average. A small percentage of hopelessly connected surfers conduct hundreds of searches a day, and many more do no more than one or two a month. (All these figures, as one might expect, are growing over time.)

The process of how we get our results starts with the crawler. The crawler is a specialized software program that hops from link to link on the World Wide Web, scarfing up the pages it finds and sending them back to be indexed. It's seductive to think of crawlers as tiny little robots wandering the vast halls of cyberspace, but the truth is a bit more mundane. Crawlers are in fact homebodies, sitting on their own servers and sending out vast numbers of requests

Those requests bring back Web pages, which the crawler then hands off to the indexer. It also takes note of any links it has found on the page, and queues those links in its request file—sending out yet more requests to the newly found links, which find more links . . . and so on, ad infinitum. Though the science behind crawlers is complex, what they do is pretty simple: they go off on a endless binge of dialing for URLs, and they report back what they've found. Crawlers have long been the least visible of the search engine's components, but they are arguably the most important. The more sites they crawl, and the more frequently they crawl them, the more complete the index is. When the index is more complete, the search results pages (SERPs) that are returned for a particular query have a greater chance of being relevant.

Early versions of crawlers discovered and indexed only the titles of Web pages, but today's more advanced versions index the contents of the entire Web page, as well as many different file types such as Adobe Acrobat (PDF), Microsoft Office documents, audio and video, and even site-specific metadata—structured information provided by site owners about the pages or information being crawled.

The crawler sends its data back to a massive database called the index. The index breaks into several pieces, depending on whether the data has been processed and made ready for consumption by searchers like you and me. Raw indexes are rather like lists organized by domain: for any given site, the index will list all the pages on that site, as well as all pertinent information about those pages: the words on the page, the links, the anchor text (text around and within a link), and so on. The information is organized in such a way that if you know the URL you can find the words that are related to that URL.

Why is this important? Because the next step in creating a smart index is to invert the database—in essence, to make a list of words that are then associated with URLs. So when you type "outer Mongolia" into a search box, the engine immediately can retrieve a list of

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.