

# FOREWORD

---

Just a decade ago, politicians were advocating a future with something called the “Information Superhighway.” But they didn't realize that we already had it: the Internet. The Internet is the mechanism that underlies many of today's tools: it enables electronic mail and electronic commerce, facilitates file sharing and scientific research, and connects the World Wide Web. Even though the Internet has obvious utility beyond the Web, it took the World Wide Web and a browser called Mosaic to bring the Internet from the research community to the world's attention.

In a matter of a few short years, the World Wide Web has become our most important information resource. Universities and schools use it for research and learning. Companies use it to streamline operations with suppliers and customers. Governments use it to communicate both across national boundaries and internally with citizens. And individuals use it to connect with family and friends scattered by modern society. The amount of information available on the Web is expanding so quickly that it's difficult to imagine.

Anyone can publish anything on the Web, so the size of the Web is growing not only as traditional publishers add more content, but also as the number of authors expands. That everyone can be an author is the unique and powerful force behind the Web. But this fact also makes finding information much more difficult: every opinion is on the Web, every side of an issue, every truth, and many falsehoods. Which one is right? How do you find what you need?

It was in this environment, nearly 10 years ago, that I created WebCrawler. At the time, the Internet was still primarily a research tool and the Web was small, consisting of perhaps 6,000 sites. WebCrawler was a tool for finding information on those sites by content: you typed what you wanted to find and WebCrawler would look in its index to find Web pages that matched your query. It was among the first search engines on the Web, and was the very first to index the full-text of every page it found. WebCrawler created the index by automatically “crawling” the Web: moving from link to link and downloading Web pages.

During 1994, the Web grew dramatically. WebCrawler had to expand with it and handle the growing number of queries from the larger number of people connected to the Web. After a year of operation, WebCrawler was handling 300,000 requests per day. After three years, WebCrawler was handling nearly eight million requests per day against an index of about a million documents.

Today, WebCrawler is just one of a large number of search engines. The Web contains billions of pages and has hundreds of millions of users. The task of creating a search engine from scratch is difficult because of the Web's size: with tens of billions of pages, which few billion should the search engine index?

It's a fact today that people make money by selling products and services via the Web. As traffic to their Web sites increases, they make more money. The enormous commercial potential of the Web means that good search engines add real value: they level the playing field by trying to keep those with commercial messages from overwhelming your search results. Just as businesses use electronic mail to drive traffic to their sites, they use placement on search engines to increase traffic. Since searchers usually click on just the first few links that search engines return, being at the top of the listings can garner one a lot of business. Web page authors employ all kinds of tricks to achieve such placement, and in doing so they create what's become the biggest problem for search engines: separating the stuff that honest people want to find from the stuff that businesses (even honest businesses) want them to find.

Search engines come in many flavors, each with its own features, quirks, and commercial bent. Knowing not only how to find an answer, but also how to find it such that you have confidence in the result is a crucial skill.

The Old Fart's Guide to Internet Searches is a delightful read and a rarity on the bookshelves. Most books on the market today are either technical or focus entirely on the history of the World Wide Web. Within the pages of this book, you will find a clear and enjoyable account of the birth of the Web and its evolution into today's most important information resource. Aaron then focuses on searching the Web: how to find exactly what you are interested in. Aaron does this with extensive case studies followed by a technical section that talks to you, not above you. All this, plus a no-nonsense description of how to find a quality Internet Service Provider, makes the Old Fart's Guide to Internet Searches an indispensable reference.

Brian Pinkerton  
Woodside, CA  
December 16, 2003