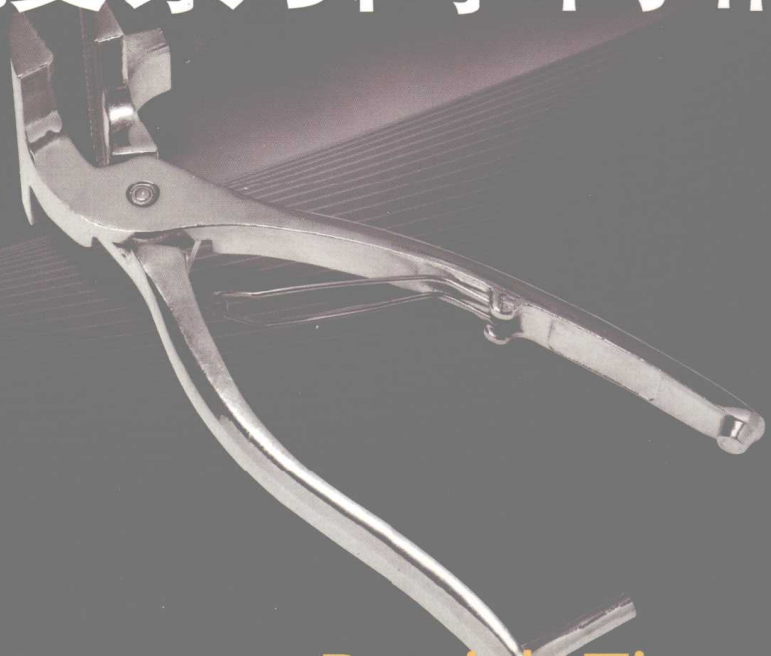
Microsoft

Inside the Index and Search Engines:

Microsoft® Office SharePoint® Server 2007

索引及搜索引擎内幕



China Edition
Published by Shanghai World Publishing Corporation

Patrick Tisseghem
Lars Fastrup

世界图书出版公司

Microsoft®

Inside the Index and Search Engines: Microsoft® Office SharePoint® Server 2007

Patrick Tisseghem
Lars Fastrup

图书在版编目(CIP)数据

Microsoft Office SharePoint Server 2007 索引及搜索引擎内幕:英文/(俄罗斯)蒂瑟汉姆,(俄罗斯)法斯特鲁普著. -上海:上海世界图书出版公司,2008.10

ISBN 978-7-5062-9290-0

Ⅰ. M… Ⅱ. ①蒂…②法… Ⅲ. 企业管理-应用软件,Microsoft Office SharePoint Server 2007-英文 Ⅳ. F270.7

中国版本图书馆 CIP 数据核字(2008)第 159890 号

Microsoft Office SharePoint Server 2007

索引及搜索引擎内幕

[俄]帕特里克·蒂瑟汉姆 拉斯·法斯特鲁普 著

上海世界图书出版公司出版发行

上海市尚文路 185 号 B 楼

邮政编码 200010

(公司电话:021-63783016 转发行部)

上海竟成印务有限公司印刷

如发现印装质量问题,请与印刷厂联系

(质检科电话:021-56422678)

各地新华书店经销

开本:787×960 1/16 印张:40 字数:1300 000

2008 年 10 月第 1 版 2008 年 10 月第 1 次印刷

ISBN 978-7-5062-9290-0/T·177

图字:09-2008-627 号

定价:238.00 元

http://www.wpcsh.com.cn

http://www.mspress.com.cn

Acknowledgments

Writing a book while you have a full-time job during the day is very stressful, not only personally, but also for the family. That's why I want to thank, above all, my wife Linda and my two lovely daughters, Laura and Anahi, for their support once again during these past eight months. Lars was one of the students in my first SharePoint 2003 developer class in the summer of 2003. Since then, we have been close friends, and the choice of him as a co-author for this book was obvious due to his expertise in this domain. I would like to thank him for all of his efforts. Without him, the book would be totally developer-focused. He made sure that there are plenty of administrator-focused topics in the book. I also would like to thank Karine and the whole U2U team for their support; the people at Microsoft Press; the editors Lori, Steve, Jennifer, Randall, and Rosemary; Mike Fitzmaurice for his continuous support (even out there in the Antartic), and all of my students who have inspired me a lot during the many trainings in different countries I have led the past 2 years.

—Patrick Tisseghem

This is my first SharePoint book project, and before anything, I would like to thank my buddy Patrick Tisseghem for offering the opportunity to co-author this book. It has been a challenging and rewarding experience working with you on this project as it takes a lot to keep up with your standard of work. But your easygoing nature is very compatible with mine and I truly like working with you. I especially enjoyed our first review session at my place in Denmark, where our shared interest in a good single malt whisky also concluded the night very nicely. My acknowledgment also goes to the people at Microsoft Press for accepting and trusting me as an author. I would foremost like to thank my primary contacts, Lori Merrick, Rosemary Caperton, Steve Sagman, Randall Galloway, and Jennifer Harris, for your patience and competent approach to reviewing my chapters. I am impressed with the quality your work has added to the final result. Finally, I would also like to thank Microsoft and the SharePoint Search team for building and delivering a product that adds great value to the market. This product has played an important role in my professional career and has helped me build a great professional network and a valuable technical skill set. I can hardly wait to see what you come up with in the next version.

—Lars Fastrup

Introduction

Over the past decade, we have seen a large investment from Microsoft in the Enterprise Search space, and there's more to come. At the time of this writing, we have support for search in Windows SharePoint Services 3.0 and in Microsoft Office SharePoint Server 2007, there is the new release of Microsoft Search Server 2008, and community tools are available for extending the search infrastructure and the search user experience. There is also a wealth of Microsoft partners who are active in the Enterprise Search space, filling the niches; selling helpful add-ons that enrich the experience for administrators, users, and developers; and delivering consultancy services around of all this.

We (Lars Fastrup and Patrick Tisseghem) decided to team up in the summer of 2007 to write this book. We were, and still are, convinced that the time is right to have a dedicated book covering all aspects of the indexing and search infrastructure that is available when you deploy Microsoft Office SharePoint Server 2007 within an organization. Lars has spent most of his professional career working in the Enterprise Search space on a highly successful third-party product named Ontolica Search. Patrick is the author of *Inside Microsoft Office SharePoint Server 2007* (Microsoft Press, 2007), he educates people about SharePoint development, and he is very passionate about the options for customizing and extending the search infrastructure with custom solution components.

Target Audience

The result is a book covering a wide range of topics for both administrators and developers. And as you'll notice while reading through these chapters, very often there is a thin line in the world of SharePoint between what administrators have to do and what developers have to do. Both must know about each others' tasks to a certain degree.

This book is of course based on a lot of information that is contained in the Microsoft Software Developer Kit (SDK) for both Windows SharePoint Services 3.0 and Office SharePoint Server 2007. It is good to have this resource available while you go through the content. An online version is available from the Microsoft Developer Network (MSDN) at *http://msdn2.microsoft.com/en-us/library/bb931736.aspx*, or you can download the full documentation at *http://www.microsoft.com/downloads/details.aspx?FamilyId=6D94E307-67D9-41AC-B2D6-0074D6286FA9*. Administrators can get more background information on the Microsoft Technet site at *http://technet.microsoft.com/en-us/library/cc263630.aspx*.

Code Samples

The language used in the book is C#, and sample code is available only in C#. All of the examples are wrapped in single Microsoft Visual Studio 2005 or Microsoft Visual Studio 2008 projects and can be downloaded from the book's companion Web site at the following address:

http://www.microsoft.com/mspress/companion/9780735625358/

Microsoft Press provides support for books and companion content at the following Web site:

http://www.microsoft.com/learning/support/books/

Questions and Comments

If you have comments or ideas regarding this book or the companion content, or if you have questions that are not answered by visiting the preceding Web sites, please send them to Microsoft Press via e-mail to

mspinput@microsoft.com

Please note that Microsoft software product support is not offered through the preceding address.

Contents at a Glance

Table of Contents

Chapter 1

Introducing Enterprise Search in SharePoint 2007

After completing this chapter, you will be able to

- Describe the role of Microsoft in the Enterprise Search space.
- Distinguish between the search features supported in the various Microsoft Enterprise Search products.
- See how Federated Search is supported in Microsoft Search Server 2008.
- Understand the index and search architecture of the Microsoft Office SharePoint Products And Technologies at a high level.
- Describe the administrative and developer-oriented topics covered in this book.

The Importance of Search and the Role of Microsoft

"Knowledge is of two kinds: we know a subject ourselves, or we know where we can find information upon it." Samuel Johnson's thinking in the eighteenth century is obviously still very relevant in the world we are living in today. The amount of information that is stored in digital format is enormous, and growing more every day. As humans, it has become almost impossible for us to be knowledgeable in the many subjects that govern our lives without some support of software tools that help us in the quest to find the right information at the right time. This is true for all the things we carry out in our private occupations, but it is even more important within organizations, where there is a growing need for sophisticated software support that enables information workers to find the required piece of corporate information as fast as possible and regardless of where that piece of data is stored.

The past decade has seen a lot of investment from Microsoft in search-related technologies, with a focus on three search spaces: the Internet, the desktop, and internal corporate networks, known collectively as the *enterprise*. On the Internet, Microsoft delivers its search experience through Live Search (*http://www.live.com*), shown in Figure 1-1. Although Live Search is facing tough competition from search engines such as Google and Yahoo, the latest figures show a growing market share. This is due to one of the strongest features of Live Search: its integration with the other Microsoft online services.

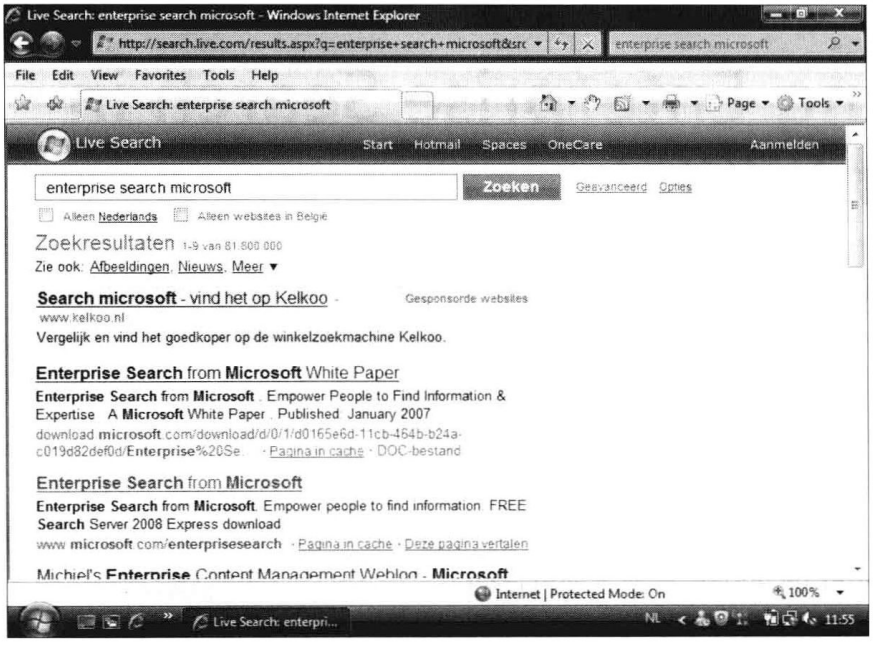


FIGURE 1-1 Microsoft Live Search

Search support in the desktop space is where Microsoft has a very strong position, with indexing and search technologies in the latest client operating systems, such as Microsoft Windows Vista, as shown in Figure 1-2.

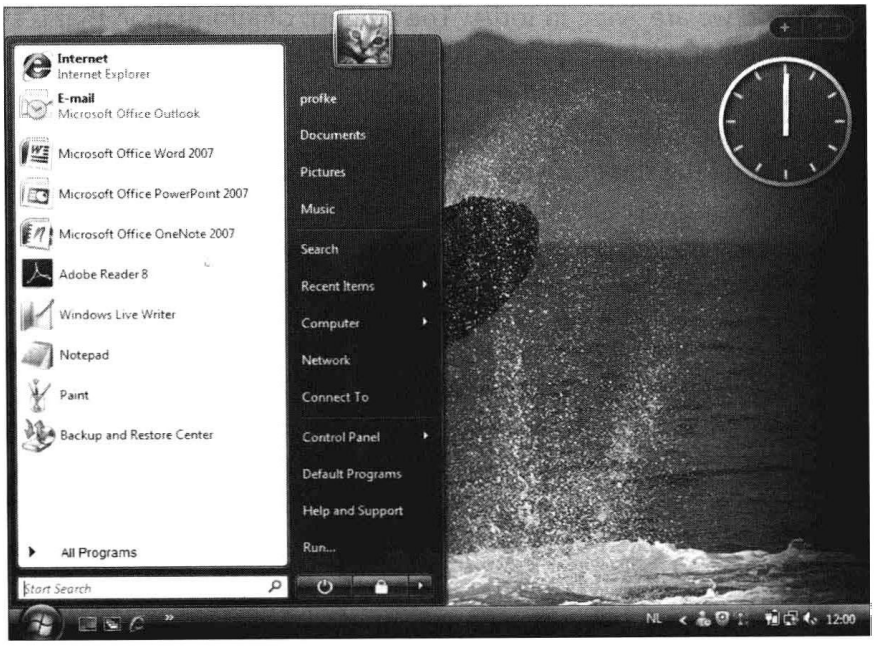


FIGURE 1-2 Desktop search from the Start menu in Windows Vista

Many users see the search capabilities of Windows Vista, displayed on the Start menu, as shown in Figure 1-2, and in Microsoft Windows Explorer, as shown in Figure 1-3, as one of the best new features of the operating system. All of the data that is stored on a laptop or PC is indexed in the background in a nondisruptive way, and when it's needed, users can quickly find the data without being too worried about the place where they stored that data. The integration of all the support for search in popular software products such as Microsoft Office Outlook 2007 is definitely beneficial for information workers who want, for example, to find information that is stored in the many e-mail messages that are typically part of their mailboxes.

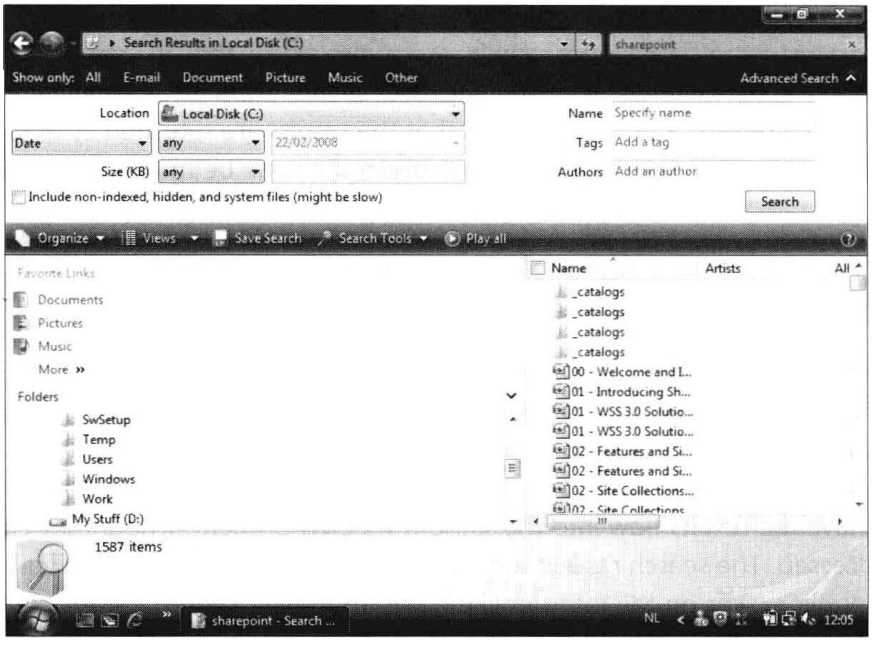


FIGURE 1-3 Searching the desktop from Windows Explorer in Windows Vista

The Internet and the desktop search space are beyond the scope of this book. Our focus will be on the support that Microsoft provides for indexing and searching of data stored within internal corporate networks. Support for searching within this area appeared more than ten years ago, with technologies such as the Windows Indexing Service, which was included in early versions of Internet Information Server (IIS) and is the precursor to the Microsoft SharePoint Products And Technologies: Microsoft Site Server 3.0. Support for search on corporate networks was initially concentrated on the indexing of data stored in database systems such as Microsoft SQL Server, but it has now been expanded to data that is stored within a multitude of various types of data storage systems.

Although support in these search spaces started as fairly separate efforts, today Microsoft is putting a lot of effort into the development and concentration of all of the search-related technologies into a single core search technology that is integrated within the majority of a company's products but customized to its own audiences. Clearly, employees should not be

limited to support only for searching data that resides within the corporate database systems or shared folders. Information workers can be empowered with support to find information that comes from many places: the outside world, the local file system, and the data stores (whatever form they have) within the company. This uniting of the search spaces into one is also known as *Enterprise Search*.

More Information You can learn more about Enterprise Search at *http://www.microsoft.com/enterprisesearch*.

Enterprise Search is not just about indexing and searching information, however. Microsoft wants to go a step further and make the search results also actionable. In short, there are three important design goals: to enhance the end-user search experience, to include people's expertise in the search results, and to make sure that the results are returned as quickly as possible, in the most relevant order. And all of this must of course be manageable for the administrators.

The Search User Experience

The search for information ought to be a natural experience for users working in Microsoft desktop and server products. Without leaving a comfort zone such as Microsoft Office Outlook, Microsoft Office Word, or a document workspace in Microsoft Windows SharePoint Services, a user should be able to enter search queries, have the queries executed behind the scenes, and have the results flow into the same work environment, where they can be immediately processed. The search results should also provide information from various locations, including public places on the Internet like blogs and Wikis, shared folders containing Microsoft Office system documents on the intranet, and documents located on an external hard disk connected to the user's laptop. In other words, the user should be able to search in many locations from a single place.

The context in which the search is executed is equally important and should be reflected in the display of the search results. A salesperson operating in the North American region should not see the same search results as a salesperson working in the European region. Most companies capture and store contextual information such as details about the work environment and the responsibililities of the employees already and make it available in some form. Search engines should be able to pick up the information and adapt the search results accordingly. Users should also be able to see only the items in the search results that they are allowed to see based on access control lists (ACLs) that are associated with the containers storing the information. In some instances, the security information is associated with the individual piece of information itself. In all cases, a trimming of the search results should take place and should be transparent to the user and to the administrators who are responsible for the management of the search infrastructure.