

## Building A Corporate Taxonomy Benefits And Challenges

"This book addresses e-learning patterns in software development, providing an accessible language to communicate sophisticated knowledge and important research methods and results"--Provided by publisher.

. . . the editors have done a good job of bringing together a series of contributions which provide a useful and welcome expansion of the theoretical foundations of talent management through a knowledge management lens. David Collings, Personnel Review This book takes a fresh look at human talent in organizations, focusing on employees at all levels who represent key agents of knowledge management in acquiring, transferring, and applying important knowledge for competitive advantage. The overarching aim of the book is to identify, define, and explore the implementation of talent management strategies aimed at facilitating effective knowledge management in an organization. The contributors provide a valuable fusion of two important areas of emphasis for current research and practice in human resource management: talent management and knowledge management. They illustrate the immense significance of the latter to competitive advantage and organizational success in our rapidly changing global knowledge-based economy. The generation and acquisition of ideas and knowledge, their internal transfer and application throughout the organization, and the cross-border transfer of knowledge all through the effective management of human talent have become integral to contemporary management. The contributors examine planning and staffing, training/coaching, performance management, and organizational learning and development. Academics, human resource management practitioners and management consultants will find this volume valuable.

Successfully architect a Drupal 8 website that scales to meet project requirements of any size and scope. Starting with a one-chapter review of Drupal basics, you'll dive into deeper topics including software development processes, architecting a Drupal site, scaling Drupal, working effectively with themes, and more. In addition to a thorough discussion of custom module development and how to develop modules as building blocks, you'll also review many common ways of integrating Drupal with other 3rd party systems. Building and maintaining an enterprise-ready Drupal website presents a unique set of challenges and complexities. From development processes and content management to deployments, version control and more, all aspects of Drupal development are impacted when two or more developers are on a project. If you are involved in a Drupal project that requires frequent updates and long-term support from a team of developers, system administrators, and end-users, Enterprise Drupal 8 Development is for you. Whether you're an experienced Drupal developer looking to expand your skills, a systems administrator managing a Drupal project, or a PHP developer new to Drupal, Enterprise Drupal 8 Development will give you the knowledge and inspiration you need to manage large and complex projects. What You'll Learn Architect Drupal for development teams and apply best practices for large Drupal projects Integrate 3rd party applications with Drupal Manage content, code, and configuration across multiple staging environments Build a consistent user experience for administrators and content creators Who This Book Is For The primary audience is

intermediate to advanced Drupal developers who are working on large-scale Drupal projects and advanced PHP developers new to Drupal The secondary audience is IT Professionals such as Systems Administrators managing Drupal projects Industry professionals, government officials, and the general public often agree that the modern healthcare system is in need of an overhaul. With many organizations concerned with the long-term care of patients, new strategies, practices, and organizational tools must be developed to optimize the current healthcare system. Healthcare Policy and Reform: Concepts, Methodologies, Tools, and Applications is a comprehensive source of academic material on the importance of policy and policy reform initiatives in modern healthcare systems. Highlighting a range of topics such as public health, effective care delivery, and health information systems, this multi-volume book is designed for medical practitioners, medical administrators, professionals, academicians, and researchers interested in all aspects of healthcare policy and reform.

Software Engineering and Computer Systems, Part II Second International Conference ICSECS 2011, Kuantan, Pahang, Malaysia, June 27-29, 2011, Proceedings Springer IBM® FileNet® Content Manager Version 5.2 provides full content lifecycle and extensive document management capabilities for digital content. IBM FileNet Content Manager is tightly integrated with the family of IBM FileNet products based on the IBM FileNet P8 technical platform. IBM FileNet Content Manager serves as the core content management, security management, and storage management engine for the products. This IBM Redbooks® publication covers the implementation best practices and recommendations for solutions that use IBM FileNet Content Manager. It introduces the functions and features of IBM FileNet Content Manager, common use cases of the product, and a design methodology that provides implementation guidance from requirements analysis through production use of the solution. We address administrative topics of an IBM FileNet Content Manager solution, including deployment, system administration and maintenance, and troubleshooting. Implementation topics include system architecture design with various options for scaling an IBM FileNet Content Manager system, capacity planning, and design of repository design logical structure, security practices, and application design. An important implementation topic is business continuity. We define business continuity, high availability, and disaster recovery concepts and describe options for those when implementing IBM FileNet Content Manager solutions. Many solutions are essentially a combination of information input (ingestion), storage, information processing, and presentation and delivery. We discuss some solution building blocks that designers can combine to build an IBM FileNet Content Manager solution. This book is intended to be used in conjunction with product manuals and online help to provide guidance to architects and designers about implementing IBM FileNet Content Manager solutions. Many of the features and practices described in the book also apply to previous versions of IBM FileNet Content Manager.

Discusses Web site hierarchy, usability, navigation systems, content labeling, configuring search systems, and managing the information architecture development process.

"This book addresses the gap in current literature in terms of linking and understanding the relationship between e-government and government

enterprise architecture"--Provided by publisher.

This book illustrates the key role played by taxonomy in the conservation and sustainable utilisation of plant biodiversity. It is a tribute to the work of Professor Vernon Heywood who has done so much to highlight the importance of sound scholarship, training and collaboration for plant conservation. Divided into four parts, the book opens with an overview of the place of taxonomy in science and in implementing the Convention on Biological Diversity. Part 2 outlines the theoretical basis of taxonomy, how it is done and how it contributes to measuring diversity. The third part explains how taxonomy is used to establish conservation priorities and actions and the concluding part illustrates taxonomy in the practice and measurement of effective conservation action. With contributions from taxonomists and also the users of taxonomy, the volume will provide a balanced treatment, suitable for advanced students, researchers and conservation professionals.

Making Enterprise Information Management (EIM) Work for Business: A Guide to Understanding Information as an Asset provides a comprehensive discussion of EIM. It endeavors to explain information asset management and place it into a pragmatic, focused, and relevant light. The book is organized into two parts. Part 1 provides the material required to sell, understand, and validate the EIM program. It explains concepts such as treating Information, Data, and Content as true assets; information management maturity; and how EIM affects organizations. It also reviews the basic process that builds and maintains an EIM program, including two case studies that provide a birds-eye view of the products of the EIM program. Part 2 deals with the methods and artifacts necessary to maintain EIM and have the business manage information. Along with overviews of Information Asset concepts and the EIM process, it discusses how to initiate an EIM program and the necessary building blocks to manage the changes to managed data and content. Organizes information modularly, so you can delve directly into the topics that you need to understand Based in reality with practical case studies and a focus on getting the job done, even when confronted with tight budgets, resistant stakeholders, and security and compliance issues Includes applicatory templates, examples, and advice for executing every step of an EIM program

This forward-looking book examines the issue of intellectual property (IP) law reform, considering both the reform of primary IP rights, and the impact of secondary rights on such reforms. It reflects on the distinction between primary and secondary rights, offering new international perspectives on IP reform, and exploring both the intended and unintended consequences of changing primary rights or adding secondary rights.

Electronic Commerce Management for Business Activities and Global Enterprises: Competitive Advantages is written as an e-commerce textbook for undergraduate and graduate students in various business programs, including information systems, marketing, computer science, and MBA. In addition to

serving as a textbook in e-commerce, this book also provides an excellent repository for instructors, researchers, and industry practitioners for their research ideas, theories, and practical experiences. In addition to regular topics traditionally taught in the classroom, this textbook addresses the many new emerging ideas and applications and presents tools and techniques in all aspects of e-commerce development and management in the global economy.

"Explains how to create terms and relationships, select taxonomy management software, design taxonomies for human versus automated indexing, manage enterprise taxonomy projects, and adapt taxonomies to various user interfaces." - cover.

The second edition of this best-selling Handbook presents a fully updated and expanded overview of research, providing the latest perspectives on the analysis of theories, techniques, and methods used by industrial, work, and organizational psychologists. Building on the strengths of the first edition, key additions to this edition include in-depth historical chapter overviews of professional contexts across the globe, along with new chapters on strategic human resource management; corporate social responsibility; diversity, stress, emotions and mindfulness in the workplace; environmental sustainability at work; aging workforces, among many others. Providing a truly global approach and authoritative overview, this three-volume Handbook is an indispensable resource and essential reading for professionals, researchers and students in the field. Volume One: Personnel Psychology and Employee Performance Volume Two: Organizational Psychology Volume Three: Managerial Psychology and Organizational Approaches

The ultimate guide to successful interviews for Enterprise, Business, Domain, Solution, and Technical Architect roles as well as IT Advisory Consultant and Software Designer roles About This Book Learn about Enterprise Architects IT strategy and NFR – this book provides you with methodologies, best practices, and frameworks to ace your interview A holistic view of key architectural skills and competencies with 500+ questions that cover 12 domains 100+ diagrams depicting scenarios, models, and methodologies designed to help you prepare for your interview Who This Book Is For This book is for aspiring enterprise, business, domain, solution, and technical architects. It is also ideal for IT advisory consultants and IT designers who wish to interview for such a role. Interviewers will be able leverage this book to make sure they hire candidates with the right competencies to meet the role requirements. What You Will Learn Learn about IT strategies, NFR, methodologies, best practices, and frameworks to ace your interview Get a holistic view of key concepts, design principles, and patterns related to evangelizing web and Java enterprise applications Discover interview preparation guidelines through case studies Use this as a reference guide for adopting best practices, standards, and design guidelines Get a better understanding with 60+ diagrams depicting various scenarios, models, and methodologies Benefit from coverage of all architecture domains including EA

(Business, Data, Infrastructure, and Application), SA, integration, NFRs, security, and SOA, with extended coverage from IT strategies to the NFR domain In Detail An architect attends multiple interviews for jobs or projects during the course of his or her career. This book is an interview resource created for designers, consultants, technical, solution, domain, enterprise, and chief architects to help them perform well in interview discussions and launch a successful career. The book begins by providing descriptions of architecture skills and competencies that cover the 12 key domains, including 350+ questions relating to these domains. The goal of this book is to cover all the core architectural domains. From an architect's perspective, it is impossible to revise or learn about all these key areas without a good reference guide – this book is the solution. It shares experiences, learning, insights, and proven methodologies that will benefit practitioners, SMEs, and aspirants in the long run. This book will help you tackle the NFR domain, which is a key aspect pertaining to architecting applications. It typically takes years to understand the core concepts, fundamentals, patterns, and principles related to architecture and designs. This book is a goldmine for the typical questions asked during an interview and will help prepare you for success! Style and approach This book will help you prepare for interviews for architectural profiles by providing likely questions, explanations, and expected answers. It is an insight-rich guide that will help you develop strategic, tactical, and operational thinking for your interview.

An enterprising Computing series title that focuses on the power of ISD to increase productivity by defining scope of services and responsibilities throughout the IT Enterprise.

This Three-Volume-Set constitutes the refereed proceedings of the Second International Conference on Software Engineering and Computer Systems, ICSECS 2011, held in Kuantan, Malaysia, in June 2011. The 190 revised full papers presented together with invited papers in the three volumes were carefully reviewed and selected from numerous submissions. The papers are organized in topical sections on software engineering; network; bioinformatics and e-health; biometrics technologies; Web engineering; neural network; parallel and distributed e-learning; ontology; image processing; information and data management; engineering; software security; graphics and multimedia; databases; algorithms; signal processing; software design/testing; e- technology; ad hoc networks; social networks; software process modeling; miscellaneous topics in software engineering and computer systems.

The continuous development of new technologies has led to significant socio-economic advances in modern society. When applied in the medical sector, healthcare delivery techniques are optimized. Health Information Systems and the Advancement of Medical Practice in Developing Countries is a comprehensive reference source for the latest scholarly research on technology utilization for delivering reliable and accurate health information to patients and clinical staff. Highlighting pivotal perspectives on topics such as mobile health,

telemedicine, and healthcare access, this book is ideally designed for professionals, practitioners, researchers, academics, and graduate students interested in the benefits and challenges of technology applications in healthcare systems.

- Corporate Web portals are increasingly common, providing employees and customers with one easy to use online access point

- Provides a high level yet practical overview of the concepts, technologies, and products used in building successful portals

- Describes the principles and methodologies for crafting and executing a successful business-aligned IT strategy to provide businesses with value delivery.

This edited collection provides a cutting edge overview of issues of key concern for information professionals providing information services in corporate environments. Corporate information professionals serving the workplace rather than learning communities or the general public face specific challenges and demands, from providing competitive intelligence to managing information in a global environment. International contributors working across a variety of sectors pinpoint the key topics facing the corporate information professionals today and share their experiences and expertise. The key topics include: - how information professionals/librarians fit into the contemporary workplace - managing the corporate intranet - the role of the corporate librarian in internal and external marketing - gaining buy-in for corporate knowledge and information management - the hybrid librarian/systems specialist - managing staff and change in a difficult climate - demonstrating value managing information in a global firm - developing corporate taxonomies at a time of change - working with suppliers/licensing for e-libraries - training end-users - competitive intelligence searching. Readership: Experienced information professionals working in the corporate sector, including professional services firms, government, NGOs, commercial and industrial companies. The book should be useful to those with a high level of experience and/or seniority, wanting an overview on specific aspects of corporate information management, but will be accessible to more recent entrants to the workplace. It will also be of interest to students of librarianship and those applying for jobs within the sector, as well as the related professions of knowledge management, information architecture and intranet management.

The ultimate guide to electronic records management, featuring a collaboration of expert practitioners including over 400 cited references documenting today's global trends, standards, and best practices Nearly all business records created today are electronic, and are increasing in number at breathtaking rates, yet most organizations do not have the policies and technologies in place to effectively organize, search, protect, preserve, and produce these records. Authored by an internationally recognized expert on e-records in collaboration with leading subject matter experts worldwide, this authoritative text addresses the widest range of in-depth e-records topics available in a single volume. Using guidance from information governance (IG) principles, the book covers methods and best practices for everything from new e-records inventorying techniques and retention schedule development, to taxonomy

design, business process improvement, managing vital records, and long term digital preservation. It goes further to include international standards and metadata considerations and then on to proven project planning, system procurement, and implementation methodologies. Managing Electronic Records is filled with current, critical information on e-records management methods, emerging best practices, and key technologies. Thoroughly introduces the fundamentals of electronic records management Explains the use of ARMA's Generally Accepted Recordkeeping Principles (GARP®) Distills e-records best practices for email, social media, and cloud computing Reveals the latest techniques for e-records inventorying and retention scheduling Covers MS SharePoint governance planning for e-records including policy guidelines Demonstrates how to optimally apply business process improvement techniques Makes clear how to implement e-document security strategies and technologies Fully presents and discusses long term digital preservation strategies and standards Managing e-records is a critical area, especially for those organizations faced with increasing regulatory compliance requirements, greater litigation demands, and tightened internal governance. Timely and relevant, Managing Electronic Records reveals step-by-step guidance for organizing, managing, protecting, and preserving electronic records.

Whilst there are many books on knowledge management there are few aimed directly at HR practitioners and the critical role that they can play in building a knowledge-centric culture. This practical book draws on the author's own experience, as well as that of leading-edge Human Resource and Knowledge Management practitioners (including Linda Holbeche, Elizabeth Lank, and David Snowden), each of whom recognise that building a knowledge-centric culture cannot be achieved through technology alone. It covers areas such as:

- \* Defining the key ingredients of a knowledge-centric culture
- \* The changing structures, roles and responsibilities needed to create a knowledge-centric culture
- \* HR's unique contribution to building a knowledge-centric culture, together with practical steps for getting started on the KM journey and for keeping the momentum going
- \* Tools and techniques for: opening up a dialogue about why knowledge management is crucial for business and personal success; knowledge mapping; encouraging and facilitating knowledge sharing, as well as ways of identifying key knowledge players
- \* How to help your organisation reframe its assumptions about learning in the knowledge economy
- \* How to ensure that your HR practices are knowledge aligned

As tech giants and startups disrupt every market, those who master large-scale software delivery will define the economic landscape of the 21st century, just as the masters of mass production defined the landscape in the 20th. Unfortunately, business and technology leaders are woefully ill-equipped to solve the problems posed by digital transformation. At the current rate of disruption, half of S&P 500 companies will be replaced in the next ten years. A new approach is needed. In Project to Product, Value Stream Network pioneer and technology business leader Dr. Mik Kersten introduces the Flow Framework—a new way of seeing, measuring, and managing software delivery. The Flow Framework will enable your company's evolution from project-oriented dinosaur to product-centric innovator that thrives in the Age of Software. If you're driving your organization's transformation at any level, this is the book for you.

The implementation of an Enterprise Portal is more of an art than a science. This is true due to the fact that each organization is different with respect to priorities, fiscal issues, culture, etc. To this end, implementing an Enterprise Portal in an organization is a highly customized process that cannot be churned out in a mass-production fashion. That being said, this book will highlight some of the core processes that are similar throughout companies; how they are implemented is where the customization part comes in.

13E 2006, the 6th in this series of IFIP conferences, marked the congregation of researchers and practitioners in the areas of e-Commerce, e-Business, and e-Government. The conference was sponsored by IFIP TC 6 in cooperation with TC 8 and TC 11. The conference provided a

forum for researchers, engineers and interested users in academia, industry, and government to discuss the latest research, cutting-edge practice and upcoming trends in the growing areas of e-Commerce, e-Business, and particularly e-Government. Sophisticated applications as well as the underlying technology that supports such applications were discussed and demonstrated. The conference attracted a wide range of participants representing a significant community of researchers and practitioners from a broad range of countries. The conference was organized along parallel tracks, each track focusing on specific aspects of current research, industry applications, and public administration.

Knowledge Management has evolved into one of the most important streams of management research, affecting organizations of all types at many different levels. The Encyclopedia of Knowledge Management, Second Edition provides a compendium of terms, definitions and explanations of concepts, processes and acronyms addressing the challenges of knowledge management. This two-volume collection covers all aspects of this critical discipline, which range from knowledge identification and representation, to the impact of Knowledge Management Systems on organizational culture, to the significant integration and cost issues being faced by Human Resources, MIS/IT, and production departments.

The highly praised resource returns with updated content, examples, and code downloads WordPress is the most popular self-hosted open source website software in use today, and the latest version makes it even simpler to operate. Packed with real-world examples for load balancing and multiusers, this esteemed resource replaces some of the more basic material with more advanced content. Every single chapter has been reworked to provide the most current and proper methods for developing and designing in WordPress. Coverage of HTML5, CSS3, and JavaScript and new chapters on custom post types, custom taxonomies, Multisite, and extensions of themes bring the content of this book completely up-to-date and provides you with all you need to deploy successful WordPress sites. Offers an overview of the WordPress system and describes what happens when a WordPress-generated web page is displayed Demonstrates extending WordPress through plugins and customizing it via themes Combines a developer view of user experience and optimization with the deployment requirements for performance, security, and measurement Includes code downloads and code samples from the book's website You'll be hard pressed to find a reliable resource with more advanced content than Professional WordPress, Second Edition!

This book discusses business architecture as a basis for aligning efforts with outcomes. It views BA as complementary to enterprise architecture, where the focus of technological initiatives and inventories is to understand and improve business organization, business direction, and business decision-making. This book provides a practical, long-term view on BA. Based on the authors' consulting experience and industrial research, the material in this book is a valuable addition to the thought processes around BA and EA. The lead author has direct and practical experience with large clients in applying APQC capability framework for undertaking multiple enterprise-wide capability assessments.

These books were written as consultation books to be used to solve problems. They are essentially analogous to medical books for individuals who decided to manage the concepts and fundamentals of things in order to manage the root causes of problems. Unicist Business Objects are adaptive systems that are designed to produce a predefined result in a process. By definition they deal with the adaptive aspects of businesses. There are five types of business objects which are: the drivers, the entropy inhibitors, the inhibitors, the catalysts and the gravitational objects.

Modern businesses generate huge volumes of accounting data on a daily basis. The



recent advancements in information technology have given organizations the ability to capture and store data in an efficient and effective manner. However, there is a widening gap between this data storage and usage of the data. Business intelligence techniques can help an organization obtain and process relevant accounting data quickly and cost efficiently. Such techniques include: query and reporting tools, online analytical processing (OLAP), statistical analysis, text mining, data mining, and visualization. Business Intelligence Techniques is a compilation of chapters written by experts in the various areas. While these chapters stand on their own, taken together they provide a comprehensive overview of how to exploit accounting data in the business environment.

Adopting more sustainable ways of managing the ocean is a global priority: protecting its health will bring benefits to all. Developing countries face specific challenges, as many depend heavily on ocean-based industries and are overly exposed to the consequences of ocean degradation.

Most books on the biotechnology industry focus on scientific and technological challenges, ignoring the entrepreneurial and managerial complexities faced bio-entrepreneurs. The Business Models for Life Science Firms aims to fill this gap by offering managers in this rapid growth industry the tools needed to design and implement an effective business model customized for the unique needs of research intensive organizations. Onetti and Zucchella begin by unpacking the often-used 'business model' term, examining key elements of business model conceptualization and offering a three tier approach with a clear separation between the business model and strategy: focus, exploring the different activities carried out by the organization; locus, evaluating where organizational activities are centered; and modus, testing the execution of the organization's activities. The business model thus defines the unique way in which a company delivers on its promise to its customers. The theory and applications adopt a global approach, offering business cases from a variety of biotech companies around the world.

The present book not only illustrates the ISMS Conference's objectives and presents the contents of the lectures, but also expands on them through further in-depth contributions; the results of academic research carried out on the Conference topic, both across disciplines and ISMS member states, are thus further illuminated. The book intends to provide teachers with a guide to possible developments in the field of military science and students with greater insights into its objectives and principles. The key question is what added value, what kind of USP military science could furnish for society, what specific support not already covered by other scientific disciplines or organizations? If it were only the science of organizing, structuring and commanding armed forces, it would not be necessary to define and establish an independent scientific discipline; the military itself could accomplish all that in the required quality by itself. What would give legitimacy to an independent discipline called Military Science, is a comprehensive (i.e. not only military-strategic) organic collection of primarily intellectual capabilities, safeguarding the secure, long-term survival of the state and its people thus contributing to the benefit of society. In order to increase the publication's academic value and to provide authors with an additional incentive to contribute in future, the book was subjected to a peer review process.

Learn essential techniques from data warehouse legend Bill Inmon on how to build the

reporting environment your business needs now! Answers for many valuable business questions hide in text. How well can your existing reporting environment extract the necessary text from email, spreadsheets, and documents, and put it in a useful format for analytics and reporting? Transforming the traditional data warehouse into an efficient unstructured data warehouse requires additional skills from the analyst, architect, designer, and developer. This book will prepare you to successfully implement an unstructured data warehouse and, through clear explanations, examples, and case studies, you will learn new techniques and tips to successfully obtain and analyze text. Master these ten objectives:

- Build an unstructured data warehouse using the 11-step approach
- Integrate text and describe it in terms of homogeneity, relevance, medium, volume, and structure
- Overcome challenges including blather, the Tower of Babel, and lack of natural relationships
- Avoid the Data Junkyard and combat the “Spider’s Web”
- Reuse techniques perfected in the traditional data warehouse and Data Warehouse 2.0, including iterative development
- Apply essential techniques for textual Extract, Transform, and Load (ETL) such as phrase recognition, stop word filtering, and synonym replacement
- Design the Document Inventory system and link unstructured text to structured data
- Leverage indexes for efficient text analysis and taxonomies for useful external categorization
- Manage large volumes of data using advanced techniques such as backward pointers
- Evaluate technology choices suitable for unstructured data processing, such as data warehouse appliances

The following outline briefly describes each chapter’s content:

- Chapter 1 defines unstructured data and explains why text is the main focus of this book. The sources for text, including documents, email, and spreadsheets, are described in terms of factors such as homogeneity, relevance, and structure.
- Chapter 2 addresses the challenges one faces when managing unstructured data. These challenges include volume, blather, the Tower of Babel, spelling, and lack of natural relationships. Learn how to avoid a data junkyard, which occurs when unstructured data is not properly integrated into the data warehouse. This chapter emphasizes the importance of storing integrated unstructured data in a relational structure. We are cautioned on both the commonality and dangers associated with text based on paper.
- Chapter 3 begins with a timeline of applications, highlighting their evolution over the decades. Eventually, powerful yet siloed applications created a “spider’s web” environment. This chapter describes how data warehouses solved many problems, including the creation of corporate data, the ability to get out of the maintenance backlog conundrum, and greater data integrity and data accessibility. There were problems, however, with the data warehouse that were addressed in Data Warehouse 2.0 (DW 2.0), such as the inevitable data lifecycle. This chapter discusses the DW 2.0 architecture, which leads into the role of the unstructured data warehouse. The unstructured data warehouse is defined and benefits are given. There are several features of the conventional data warehouse that can be leveraged for the unstructured data warehouse, including ETL processing, textual integration, and iterative development.
- Chapter 4 focuses on the heart of the unstructured data warehouse: Textual Extract, Transform, and Load (ETL). This chapter has separate sections on extracting text, transforming text, and loading text. The chapter emphasizes the issues around source data. There are a wide variety of sources, and each of the sources has its own set of considerations. Extracting pointers are provided, such as reading documents only once and recognizing common and different file types.

Transforming text requires addressing many considerations discussed in this chapter, including phrase recognition, stop word filtering, and synonym replacement. Loading text is the final step. There are important points to understand here, too, that are explained in this chapter, such as the importance of the thematic approach and knowing how to handle large volumes of data. Two ETL examples are provided, one on email and one on spreadsheets. • Chapter 5 describes the 11 steps required to develop the unstructured data warehouse. The methodology explained in this chapter is a combination of both traditional system development lifecycle and spiral approaches. • Chapter 6 describes how to inventory documents for maximum analysis value, as well as link the unstructured text to structured data for even greater value. The Document Inventory is discussed, which is similar to a library card catalog used for organizing corporate documents. This chapter explores ways of linking unstructured text to structured data. The emphasis is on taking unstructured data and reducing it into a form of data that is structured. Related concepts to linking, such as probabilistic linkages and dynamic linkages, are discussed. • Chapter 7 goes through each of the different types of indexes necessary to make text analysis efficient. Indexes range from simple indexes, which are fast to create and are good if the analyst really knows what needs to be analyzed before the indexing process begins, to complex combined indexes, which can be made up of any and all of the other kinds of indexes. • Chapter 8 explains taxonomies and how they can be used within the unstructured data warehouse. Both simple and complicated taxonomies are discussed. Techniques to help the reader leverage taxonomies, including using preferred taxonomies, external categorization, and cluster analysis are described. Real world problems are raised, including the possibilities of encountering hierarchies, multiple types, and recursion. The chapter ends with a discussion comparing a taxonomy with a data model. • Chapter 9 explains ways of coping with large amounts of unstructured data. Techniques such as keeping the unstructured data at its source and using backward pointers are discussed. The chapter explains why iterative development is so important. Ways of reducing the amount of data are presented, including screening and removing extraneous data, as well as parallelizing the workload. • Chapter 10 focuses on challenges and some technology choices that are suitable for unstructured data processing. The traditional data warehouse processing technology is reviewed. In addition, the data warehouse appliance is discussed. • Chapters 11, 12, and 13 put all of the previously discussed techniques and approaches in context through three case studies: the Ablatz Medical Group, the Eastern Hills Oil Company, and the Amber Oil Company.

[Copyright: 5aef84107d30ae92fd73a988e204d471](https://www.pdfdrive.com/building-a-corporate-taxonomy-benefits-and-challenges-ebook.html)