Visual Information Retrieval Using Java And Lire Synthesis Lectures On Information Concepts Retrieval And S

The Catalan Association for Artificial Intelligence (ACIA) was formed in 1994 with the aim of promoting cooperation between researchers in artificial intelligence within the Catalan speaking community. This objective has been achieved and widened since the association held their first conference in 1998, and the annual conference of the association has become an international event presenting and discussing the latest research in AI, which attracts AI researchers from around the world. This book presents the proceedings of the 19th International Conference of the Catalan Association for Artificial Intelligence (CCIA 2016), held in Barcelona, Spain, on 19-21 October. From a total of 50 original contributions, 16 long papers and 22 short papers were accepted for presentation at the conference on the basis of their relevance, originality and technical validity. The book is divided into 7 sections: Invited Talks (synopsis only); Vision and Robotics; Logic, Constraint Satisfaction and Qualitative Theory; Classification and Clustering; Modelling; Planning and Recommender Systems; Lexical Knowledge Representation and Natural Language Processing. Providing an overview of the latest developments in the field, this book will be of interest to all those whose work involves research into, and the application of, artificial intelligence.

Everybody knows what relevance is. It is a "ya'know" notion, concept, idea–no need to explain

whatsoever. Searching for relevant information using information technology (IT) became a ubiquitous activity in contemporary information society. Relevant information means information that pertains to the matter or problem at hand—it is directly connected with effective communication. The purpose of this book is to trace the evolution and with it the history of thinking and research on relevance in information science and related fields from the human point of view. The objective is to synthesize what we have learned about relevance in several decades of investigation about the notion in information science. This book deals with how people deal with relevance—it does not cover how systems deal with relevance; it does not deal with algorithms. Spurred by advances in information retrieval (IR) and information systems of various kinds in handling of relevance, a number of basic questions are raised: But what is relevance to start with? What are some of its properties and manifestations? How do people treat relevance? What affects relevance assessments? What are the effects of inconsistent human relevance judgments on tests of relative performance of different IR algorithms or approaches? These general questions are discussed in detail.

This book constitutes the thoroughly refereed post-conference proceedings of the 17th International Conference on Information Security and Cryptology, ICISC 2014, held in Seoul, South Korea in December 2014. The 27 revised full papers presented were carefully selected from 91 submissions during two rounds of reviewing. The papers provide the latest results in research, development and applications in the field of information security and cryptology. They are organized in topical sections on RSA security, digital signature, public key cryptography, block ciphers, network security, mobile security, hash functions, information hiding and efficiency, cryptographic protocol, and side-channel attacks.

This is the second book based on the 5S (Societies, Scenarios, Spaces, Structures, Streams) approach to digital libraries (DLs). Leveraging the first volume, on Theoretical Foundations, we focus on the key issues of evaluation and integration. These cross-cutting issues serve as a bridge for those interested in DLs, connecting the introduction and formal discussion in the first book, with the coverage of key technologies in the third book, and of illustrative applications in the fourth book. These two topics have central importance in the DL field, allowing it to be treated scientifically as well as practically. In the scholarly world, we only really understand something if we know how to measure and evaluate it. In the Internet era of distributed information systems, we only can be practical at scale if we integrate across both systems and their associated content. Evaluation of DLs must take place atmultiple levels,so we can address the different entities and their associated measures. Thus, for digital objects, we assess accessibility, pertinence, preservability, relevance, significance, similarity, and timeliness. Other measures are specific to higher-level constructs like metadata, collections, catalogs, repositories, and services.We tie these together through a case study of the 5SQual tool, which we designed and implemented to perform an automatic quantitative evaluation of DLs. Thus, across the Information Life Cycle, we describe metrics and software useful to assess the quality of DLs, and demonstrate utility with regard to representative application areas: archaeology and education. Though integration has been a challenge since the earliest work on DLs, we provide the first comprehensive 5S-based formal description of the DL integration problem, cast in the context of related work. Since archaeology is a fundamentally distributed enterprise, we describe ETANADL, for integrating Near Eastern Archeology sites and information. Thus, we show how 5S-based modeling can lead to integrated services and

content. While the first book adopts a minimalist and formal approach to DLs, and provides a systematic and functional method to design and implement DL exploring services, here we broaden to practical DLs with richer metamodels, demonstrating the power of 5S for integration and evaluation.

This book introduces fundamentals of information communication. At first, concepts and characteristics of information and information communication are summarized. And then five classic models of information communication are introduced. The mechanisms and fundamental laws of the information transmission process are also discussed. In order to realize information communication, impediments in information communication process are identified and analyzed. For the purpose of investigating implications of Internet information communication, patterns and characteristics of information communication in the Internet and Web 2.0 environment are also analyzed. In the end, case studies are provided for readers to understand the theory.

While great strides have been made in the field of search and recommendation, there are still challenges and opportunities to address information access issues that involve solving tasks and accomplishing goals for a wide variety of users. Specifically, we lack intelligent systems that can detect not only the request an individual is making (what), but also understand and utilize the intention (why) and strategies (how) while providing information and enabling task completion. Many scholars in the fields of information retrieval, recommender systems, productivity (especially in task management and time management), and artificial intelligence have recognized the importance of extracting and understanding people's tasks and the intentions behind performing those tasks in order to serve them better. However, we are still

struggling to support them in task completion, e.g., in search and assistance, and it has been challenging to move beyond single-query or single-turn interactions. The proliferation of intelligent agents has unlocked new modalities for interacting with information, but these agents will need to be able to work understanding current and future contexts and assist users at task level. This book will focus on task intelligence in the context of search and recommendation. Chapter 1 introduces readers to the issues of detecting, understanding, and using task and task-related information in an information episode (with or without active searching). This is followed by presenting several prominent ideas and frameworks about how tasks are conceptualized and represented in Chapter 2. In Chapter 3, the narrative moves to showing how task type relates to user behaviors and search intentions. A task can be explicitly expressed in some cases, such as in a to-do application, but often it is unexpressed. Chapter 4 covers these two scenarios with several related works and case studies. Chapter 5 shows how task knowledge and task models can contribute to addressing emerging retrieval and recommendation problems. Chapter 6 covers evaluation methodologies and metrics for task-based systems, with relevant case studies to demonstrate their uses. Finally, the book concludes in Chapter 7, with ideas for future directions in this important research area.
With the rapid development of mobile Internet and smart personal devices in recent years, mobile search has gradually emerged as a key method with which users seek online information. In addition, cross-device search also has been regarded recently as an important research topic. As more mobile applications (APPs) integrate search functions, a user's mobile search behavior on different APPs becomes more significant. This book provides a systematic review of current mobile search analysis and studies user mobile search behavior from several

perspectives, including mobile search context, APP usage, and different devices. Two different user experiments to collect user behavior data were conducted. Then, through the data from user mobile phone usage logs in natural settings, we analyze the mobile search strategies employed and offer a context-based mobile search task collection, which then can be used to evaluate the mobile search engine. In addition, we combine mobile search with APP usage to give more in-depth analysis, such as APP transition in mobile search and follow-up actions triggered by mobile search. The study, combining the mobile search with APP usage, can contribute to the interaction design of APPs, such as the search recommendation and APP recommendation. Addressing the phenomenon of users owning more smart devices today than ever before, we focus on user cross device search behavior. We model the information preparation behavior and information resumption behavior in cross-device search and evaluate the search performance in cross-device search. Research on mobile search behaviors across different devices can help to understand online user information behavior comprehensively and help users resume their search tasks on different devices.

Research on multiculturalism and information and communication technology (ICT) has been important to understanding recent history, planning for future large-scale initiatives, and understanding unrealized expectations for social and technological change. This interdisciplinary area of research has examined interactions between ICT and culture at the group and society levels. However, there is debate within the literature as to the nature of the relationship between culture and technology. In this synthesis, we suggest that the tensions result from the competing ideologies that drive researchers, allowing us to conceptualize the relationship between culture and ICT under three primary models, each with its own

assumptions: 1) Social informatics, 2) Social determinism, and 3) Technological determinism. Social informatics views the relationship to be one of sociotechnical interaction, in which culture and ICTs affect each other mutually and iteratively, rather than linearly; the vast majority of the literature approach the relationships between ICT and culture under the assumptions of social informatics. From a socially deterministic perspective, ICTs are viewed as the dependent variable in the equation, whereas, from a technologically deterministic perspective, ICTs are an independent variable. The issues of multiculturalism and ICTs attracted much scholarly attention and have been explored under a myriad of contexts, with substantial literature on global development, social and political issues, business and public administration as well as education and scholarly collaboration. We synthesize here research in the areas of global development, social and political issues, and business collaboration. Finally we conclude by proposing under-explored areas for future research directions.

Many data-intensive applications that use machine learning or artificial intelligence techniques depend on humans providing the initial dataset, enabling algorithms to process the rest or for other humans to evaluate the performance of such algorithms. Not only can labeled data for training and evaluation be collected faster, cheaper, and easier than ever before, but we now see the emergence of hybrid human-machine software that combines computations performed by humans and machines in conjunction. There are, however, real-world practical issues with the adoption of human computation and crowdsourcing. Building systems and data processing pipelines that require crowd computing remains difficult. In this book, we present practical considerations for designing and implementing tasks that require the use of humans and machines in combination with the goal of producing high-quality labels.

Visual information retrieval (VIR) is an active and vibrant research area, which attempts at providing means for organizing, indexing, annotating, and retrieving visual information (images and videos) from large, unstructured repositories. The goal of VIR is to retrieve matches ranked by their relevance to a given query, which is often expressed as an example image and/or a series of keywords. During its early years (1995-2000), the research efforts were dominated by content-based approaches contributed primarily by the image and video processing community. During the past decade, it was widely recognized that the challenges imposed by the lack of coincidence between an image's visual contents and its semantic interpretation, also known as semantic gap, required a clever use of textual metadata (in addition to information extracted from the image's pixel contents) to make image and video retrieval solutions efficient and effective. The need to bridge (or at least narrow) the semantic gap has been one of the driving forces behind current VIR research. Additionally, other related research problems and market opportunities have started to emerge, offering a broad range of exciting problems for computer scientists and engineers to work on. In this introductory book, we focus on a subset of VIR problems where the media consists of images, and the indexing and retrieval methods are based on the pixel contents of those images -- an approach known as content-based image

retrieval (CBIR). We present an implementation-oriented overview of CBIR concepts, techniques, algorithms, and figures of merit. Most chapters are supported by examples written in Java, using Lucene (an open-source Java-based indexing and search implementation) and LIRE (Lucene Image REtrieval), an open-source Java-based library for CBIR. Table of Contents: Introduction / Information Retrieval: Selected Concepts and Techniques / Visual Features / Indexing Visual Features / LIRE: An Extensible Java CBIR Library / Concluding Remarks

Institutions typically treat research integrity violations as black and white, right or wrong. The result is that the wide range of grayscale nuances that separate accident, carelessness and bad practice from deliberate fraud and malpractice often get lost. This lecture looks at how to quantify the grayscale range in three kinds of research integrity violations: plagiarism, data falsification, and image manipulation. Quantification works best with plagiarism, because the essential one-to one matching algorithms are well known and established tools for detecting when matches exist. Questions remain, however, how many matching words of what kind in what location in which discipline constitute reasonable suspicion of fraudulent intent. Different disciplines take different perspectives on quantity and location. Quantification is harder with data falsification, because the

original data are often not available, and because experimental replication remains surprisingly difficult. The same is true with image manipulation, where tools exist for detecting certain kinds of manipulations, but where the tools are also easily defeated. This lecture looks at how to prevent violations of research integrity from a pragmatic viewpoint, and at what steps can institutions and publishers take to discourage problems beyond the usual ethical admonitions. There are no simple answers, but two measures can help: the systematic use of detection tools and requiring original data and images. These alone do not suffice, but they represent a start. The scholarly community needs a better awareness of the complexity of research integrity decisions. Only an open and wide-spread international discussion can bring about a consensus on where the boundary lines are and when grayscale problems shade into black. One goal of this work is to move that discussion forward.

Nowadays, fashion has become an essential aspect of people's daily life. As each outfit usually comprises several complementary items, such as a top, bottom, shoes, and accessories, a proper outfit largely relies on the harmonious matching of these items. Nevertheless, not everyone is good at outfit composition, especially those who have a poor fashion aesthetic. Fortunately, in recent years the number of online fashion-oriented communities, like IQON and

Chictopia, as well as e-commerce sites, like Amazon and eBay, has grown. The tremendous amount of real-world data regarding people's various fashion behaviors has opened a door to automatic clothing matching. Despite its significant value, compatibility modeling for clothing matching that assesses the compatibility score for a given set of (equal or more than two) fashion items, e.g., a blouse and a skirt, yields tough challenges: (a) the absence of comprehensive benchmark; (b) comprehensive compatibility modeling with the multi-modal feature variables is largely untapped; (c) how to utilize the domain knowledge to guide the machine learning; (d) how to enhance the interpretability of the compatibility modeling; and (e) how to model the user factor in the personalized compatibility modeling. These challenges have been largely unexplored to date. In this book, we shed light on several state-of-the-art theories on compatibility modeling. In particular, to facilitate the research, we first build three large-scale benchmark datasets from different online fashion websites, including IQON and Amazon. We then introduce a general data-driven compatibility modeling scheme based on advanced neural networks. To make use of the abundant fashion domain knowledge, i.e., clothing matching rules, we next present a novel knowledge guided compatibility modeling framework. Thereafter, to enhance the model interpretability, we put forward a prototype wise interpretable compatibility

modeling approach. Following that, noticing the subjective aesthetics of users, we extend the general compatibility modeling to the personalized version. Moreover, we further study the real-world problem of personalized capsule wardrobe creation, aiming to generate a minimum collection of garments that is both compatible and suitable for the user. Finally, we conclude the book and present future research directions, such as the generative compatibility modeling, virtual try-on with arbitrary poses, and clothing generation.

The field of human information behavior runs the gamut of processes from the realization of a need or gap in understanding, to the search for information from one or more sources to fill that gap, to the use of that information to complete a task at hand or to satisfy a curiosity, as well as other behaviors such as avoiding information or finding information serendipitously. Designers of mechanisms, tools, and computer-based systems to facilitate this seeking and search process often lack a full knowledge of the context surrounding the search. This context may vary depending on the job or role of the person; individual characteristics such as personality, domain knowledge, age, gender, perception of self, etc.; the task at hand; the source and the channel and their degree of accessibility and usability; and the relationship that the seeker shares with the source. Yet researchers have yet to agree on what context really means. While there have

been various research studies incorporating context, and biennial conferences on context in information behavior, there lacks a clear definition of what context is, what its boundaries are, and what elements and variables comprise context. In this book, we look at the many definitions of and the theoretical and empirical studies on context, and I attempt to map the conceptual space of context in information behavior. I propose theoretical frameworks to map the boundaries, elements, and variables of context. I then discuss how to incorporate these frameworks and variables in the design of research studies on context. We then arrive at a unified definition of context. This book should provide designers of search systems a better understanding of context as they seek to meet the needs and demands of information seekers. It will be an important resource for researchers in Library and Information Science, especially doctoral students looking for one resource that covers an exhaustive range of the most current literature related to context, the best selection of classics, and a synthesis of these into theoretical frameworks and a unified definition. The book should help to move forward research in the field by clarifying the elements, variables, and views that are pertinent. In particular, the list of elements to be considered, and the variables associated with each element will be extremely useful to researchers wanting to include the influences of context in their studies.

As digital collections continue to grow, the underlying technologies to serve up content also continue to expand and develop. As such, new challenges are presented whichcontinue to test ethical ideologies in everyday environs of the practitioner. There are currently no solid guidelines or overarching codes of ethics to address such issues. The digitization of modern archival collections, in particular, presents interesting conundrums when factors of privacy are weighed and reviewed in both small and mass digitization initiatives. Ethical decision making needs to be present at the onset of project planning in digital projects of all sizes, and we also need to identify the role and responsibility of the practitioner to make more virtuous decisions on behalf of those with no voice or awareness of potential privacy breaches. In this book, notions of what constitutes private information are discussed, as is the potential presence of such information in both analog and digital collections. This book lays groundwork to introduce the topic of privacy within digital collections by providing some examples from documented real-world scenarios and making recommendations for future research. A discussion of the notion privacy as concept will be included, as well as some historical perspective (with perhaps one the most cited work on this topic, for example, Warren and Brandeis' "Right to Privacy," 1890). Concepts from the The Right to Be Forgotten case in 2014 (Google Spain SL, Google Inc. v

Agencia Española de Protección de Datos, Mario Costeja González) are discussed as to how some lessons may be drawn from the response in Europe and also how European data privacy laws have been applied. The European ideologies are contrasted with the Right to Free Speech in the First Amendment in the U.S., highlighting the complexities in setting guidelines and practices revolving around privacy issues when applied to real life scenarios. Two ethical theories are explored: Consequentialism and Deontological. Finally, ethical decision making models will also be applied to our framework of digital collections. Three case studies are presented to illustrate how privacy can be defined within digital collections in some real-world examples.

Rapid technological changes and availability of news anywhere and at any moment have changed how people seek out news. Increasingly, consumers no longer take deliberate actions to read the news, instead stumbling upon news online. While the emergence of serendipitous news discovery online has been recognized in the literature, there is a limited understanding about how people experience this behavior. Based on the mixed method study that investigated online news reading behavior of residents in a Midwestern U.S. town, we explore how people accidentally discover news when engaged in various online activities. Employing the grounded theory approach, we define Incidental Exposure to

Online News (IEON) as individual's memorable experiences of chance encounters with interesting, useful, or surprising news while using the Internet for news browsing or for non-news-related online activities, such as checking email or visiting social networking sites. The book presents a conceptual framework of IEON that advances research and an understanding of serendipitous news discovery from people's holistic experiences of news consumption in their everyday lives. The proposed IEON Process Model identifies key steps in an IEON experience that could help news reporters and developers of online news platforms create innovative storytelling and design strategies to catch consumers' attention during their online activities. Finally, this book raises important methodological questions for further investigation: how should serendipitous news discovery be studied, measured, and observed, and what are the essential elements that differentiate this behavior from other types of online news consumption and information behaviors?

With the proliferation of social network services, more and more social users, such as individuals and organizations, are simultaneously involved in multiple social networks for various purposes. In fact, multiple social networks characterize the same social users from different perspectives, and their contexts are usually consistent or complementary rather than independent. Hence, as

compared to using information from a single social network, appropriate aggregation of multiple social networks offers us a better way to comprehensively understand the given social users. Learning across multiple social networks brings opportunities to new services and applications as well as new insights on user online behaviors, yet it raises tough challenges: (1) How can we map different social network accounts to the same social users? (2) How can we complete the item-wise and block-wise missing data? (3) How can we leverage the relatedness among sources to strengthen the learning performance? And (4) How can we jointly model the dual-heterogeneities: multiple tasks exist for the given application and each task has various features from multiple sources? These questions have been largely unexplored to date. We noticed this timely opportunity, and in this book we present some state-of-the-art theories and novel practical applications on aggregation of multiple social networks. In particular, we first introduce multi-source dataset construction. We then introduce how to effectively and efficiently complete the item-wise and block-wise missing data, which are caused by the inactive social users in some social networks. We next detail the proposed multi-source mono-task learning model and its application in volunteerism tendency prediction. As a counterpart, we also present a mono-source multi-task learning model and apply it to user interest inference. We

seamlessly unify these models with the so-called multi-source multi-task learning, and demonstrate several application scenarios, such as occupation prediction. Finally, we conclude the book and figure out the future research directions in multiple social network learning, including the privacy issues and source complementarity modeling. This is preliminary research on learning from multiple social networks, and we hope it can inspire more active researchers to work on this exciting area. If we have seen further it is by standing on the shoulders of giants.

In recent years there has been an increasing demand for research evaluation within universities and other research-based organisations. In parallel, there has been an increasing recognition that traditional citation-based indicators are not able to reflect the societal impacts of research and are slow to appear. This has led to the creation of new indicators for different types of research impact as well as timelier indicators, mainly derived from the Web. These indicators have been called altmetrics, webometrics or just web metrics. This book describes and evaluates a range of web indicators for aspects of societal or scholarly impact, discusses the theory and practice of using and evaluating web indicators for research assessment and outlines practical strategies for obtaining many web indicators. In addition to describing impact indicators for traditional scholarly

outputs, such as journal articles and monographs, it also covers indicators for videos, datasets, software and other non-standard scholarly outputs. The book describes strategies to analyse web indicators for individual publications as well as to compare the impacts of groups of publications. The practical part of the book includes descriptions of how to use the free software Webometric Analyst to gather and analyse web data. This book is written for information science undergraduate and Master's students that are learning about alternative indicators or scientometrics as well as Ph.D. students and other researchers and practitioners using indicators to help assess research impact or to study scholarly communication.

Chance, luck, and good fortune are the usual go-to descriptors of serendipity, a phenomenon aptly often coupled with famous anecdotes of accidental discoveries in engineering and science in modern history such as penicillin, Teflon, and Post-it notes. Serendipity, however, is evident in many fields of research, in organizations, in everyday life—and there is more to it than luck implies. While the phenomenon is strongly associated with in person interactions with people, places, and things, most attention of late has focused on its preservation and facilitation within digital information environments. Serendipity's association with unexpected, positive user experiences and outcomes has

spurred an interest in understanding both how current digital information environments support serendipity and how novel approaches may be developed to facilitate it. Research has sought to understand serendipity, how it is manifested in people's personality traits and behaviors, how it may be facilitated in digital information environments such as mobile applications, and its impacts on an individual, an organizational, and a wider level. Because serendipity is expressed and understood in different ways in different contexts, multiple methods have been used to study the phenomenon and evaluate digital information environments that may support it. This volume brings together different disciplinary perspectives and examines the motivations for studying serendipity, the various ways in which serendipity has been approached in the research, methodological approaches to build theory, and how it may be facilitated. Finally, a roadmap for serendipity research is drawn by integrating key points from this volume to produce a framework for the examination of serendipity in digital information environments.

Visual information retrieval (VIR) is an active and vibrant research area, which attempts at providing means for organizing, indexing, annotating, and retrieving visual information (images and videos) from large, unstructured repositories. The goal of VIR is to retrieve matches ranked by their relevance to a given query,

which is often expressed as an example image and/or a series of keywords. During its early years (1995-2000), the research efforts were dominated by content-based approaches contributed primarily by the image and video processing community. During the past decade, it was widely recognized that the challenges imposed by the lack of coincidence between an image's visual contents and its semantic interpretation, also known as semantic gap, required a clever use of textual metadata (in addition to information extracted from the image's pixel contents) to make image and video retrieval solutions efficient and effective. The need to bridge (or at least narrow) the semantic gap has been one of the driving forces behind current VIR research. Additionally, other related research problems and market opportunities have started to emerge, offering a broad range of exciting problems for computer scientists and engineers to work on. In this introductory book, we focus on a subset of VIR problems where the media consists of images, and the indexing and retrieval methods are based on the pixel contents of those images -- an approach known as content-based image retrieval (CBIR). We present an implementation-oriented overview of CBIR concepts, techniques, algorithms, and figures of merit. Most chapters are supported by examples written in Java, using Lucene (an open-source Java-based indexing and search implementation) and LIRE (Lucene Image REtrieval),

an open-source Java-based library for CBIR.

This book focuses on the methodologies, organization, and communication of digital image collection research that utilizes social media content. ("Image" is here understood as a cultural, conventional, and commercial—stock photo—representation.) The lecture offers expert views that provide different interpretations of images and their potential implementations. Linguistic and semiotic methodologies as well as eye-tracking research are employed to both analyze images and comprehend how humans consider them, including which salient features generally attract viewers' attention. This literature review covers image—specifically photographic—research since 2005, when major social media platforms emerged. A citation analysis includes an overview of co-citation maps that demonstrate the nexus of image research literature and the journals in which they appear. Eye tracking tests whether scholarly templates focus on the proper features of an image, such as people, objects, time, etc., and if a prescribed theme affects the eye movements of the observer. The results may point to renewed requirements for building image search engines. As it stands, image management already requires new algorithms and a new understanding that involves text recognition and very large database processing. The aim of this book is to present different image research areas and demonstrate the

challenges image research faces. The book's scope is, by necessity, far from comprehensive, since the field of digital image research does not cover fake news, image manipulation, mobile photos, etc.; these issues are very complex and need a publication of their own. This book should primarily be useful for students in library and information science, psychology, and computer science. This book seeks to advance our understanding of the relationship between information and human values by synthesizing the complementary but typically disconnected threads in the literature, reflecting on my 15 years of research on the relationship between information and human values, advancing our intellectual understanding of the key facets of this topic, and encouraging further research to continue exploring this important and timely research topic. The book begins with an explanation of what human values are and why they are important. Next, three distinct literatures on values, information, and technology are analyzed and synthesized, including the social psychology literature on human values, the information studies literature on the core values of librarianship, and the human-computer interaction literature on value-sensitive design. After that, three detailed case studies are presented based on reflections on a wide range of research studies. The first case study focuses on the role of human values in the design and use of educational simulations. The second case

study focuses on the role of human values in the design and use of computational models. The final case study explores human values in communication via, about, or using information technology. The book concludes by laying out a values and design cycle for studying values in information and presenting an agenda for further research.

With the rapid growth of web search in recent years the problem of modeling its users has started to attract more and more attention of the information retrieval community. This has several motivations. By building a model of user behavior we are essentially developing a better understanding of a user, which ultimately helps us to deliver a better search experience. A model of user behavior can also be used as a predictive device for non-observed items such as document relevance, which makes it useful for improving search result ranking. Finally, in many situations experimenting with real users is just infeasible and hence user simulations based on accurate models play an essential role in understanding the implications of algorithmic changes to search engine results or presentation changes to the search engine result page. In this survey we summarize advances in modeling user click behavior on a web search engine result page. We present simple click models as well as more complex models aimed at capturing non-trivial user behavior patterns on modern search engine result

pages. We discuss how these models compare to each other, what challenges they have, and what ways there are to address these challenges. We also study the problem of evaluating click models and discuss the main applications of click models.

With its theme, "Our Information, Always and Forever," Part I of this book covers the basics of personal information management (PIM) including six essential activities of PIM and six (different) ways in which information can be personal to us. Part I then goes on to explore key issues that arise in the "great migration" of our information onto the Web and into a myriad of mobile devices. Part 2 provides a more focused look at technologies for managing information that promise to profoundly alter our practices of PIM and, through these practices, the way we lead our lives. Part 2 is in five chapters: - Chapter 5. Technologies of Input and Output. Technologies in support of gesture, touch, voice, and even eye movements combine to support a more natural user interface (NUI). Technologies of output include glasses and "watch" watches. Output will also increasingly be animated with options to "zoom". - Chapter 6. Technologies to Save Our Information. We can opt for "life logs" to record our experiences with increasing fidelity. What will we use these logs for? And what isn't recorded that should be? - Chapter 7. Technologies to Search Our Information. The potential

for personalized search is enormous and mostly yet to be realized. Persistent searches, situated in our information landscape, will allow us to maintain a diversity of projects and areas of interest without a need to continually switch from one to another to handle incoming information. - Chapter 8. Technologies to Structure Our Information. Structure is key if we are to keep, find, and make effective use of our information. But how best to structure? And how best to share structured information between the applications we use, with other people, and also with ourselves over time? What lessons can we draw from the failures and successes in web-based efforts to share structure? - Chapter 9. PIM Transformed and Transforming: Stories from the Past, Present and Future. Part 2 concludes with a comparison between Licklider's world of information in 1957 and our own world of information today. And then we consider what the world of information is likely to look like in 2057. Licklider estimated that he spent 85% of his "thinking time" in activities that were clerical and mechanical and might (someday) be delegated to the computer. What percentage of our own time is spent with the clerical and mechanical? What about in 2057?

The information age has led to an explosion in the amount of information available to the individual and the means by which it is accessed, stored, viewed, and transferred. In particular, the growth of the internet has led to the creation of

huge repositories of multimedia documents in a diverse range of scientific and professional fields, as well as the tools to extract useful knowledge from them. Mining Multimedia Documents is a must-read for researchers, practitioners, and students working at the intersection of data mining and multimedia applications. It investigates various techniques related to mining multimedia documents based on text, image, and video features. It provides an insight into the open research problems benefitting advanced undergraduates, graduate students, researchers, scientists and practitioners in the fields of medicine, biology, production, education, government, national security and economics.

Question answering (QA) systems on the Web try to provide crisp answers to information needs posed in natural language, replacing the traditional ranked list of documents. QA, posing a multitude of research challenges, has emerged as one of the most actively investigated topics in information retrieval, natural language processing, and the artificial intelligence communities today. The flip side of such diverse and active interest is that publications are highly fragmented across several venues in the above communities, making it very difficult for new entrants to the field to get a good overview of the topic. Through this book, we make an attempt towards mitigating the above problem by providing an overview of the state-of-the-art in question answering. We cover the twin paradigms of

curated Web sources used in QA tasks ? trusted text collections like Wikipedia, and objective information distilled into large-scale knowledge bases. We discuss distinct methodologies that have been applied to solve the QA problem in both these paradigms, using instantiations of recent systems for illustration. We begin with an overview of the problem setup and evaluation, cover notable sub-topics like open-domain, multi-hop, and conversational QA in depth, and conclude with key insights and emerging topics. We believe that this resource is a valuable contribution towards a unified view on QA, helping graduate students and researchers planning to work on this topic in the near future.

Simulated test collections may find application in situations where real datasets cannot easily be accessed due to confidentiality concerns or practical inconvenience. They can potentially support Information Retrieval (IR) experimentation, tuning, validation, performance prediction, and hardware sizing. Naturally, the accuracy and usefulness of results obtained from a simulation depend upon the fidelity and generality of the models which underpin it. The fidelity of emulation of a real corpus is likely to be limited by the requirement that confidential information in the real corpus should not be able to be extracted from the emulated version. We present a range of methods exploring trade-offs between emulation fidelity and degree of preservation of privacy. We present

three different simple types of text generator which work at a micro level: Markov models, neural net models, and substitution ciphers. We also describe macro level methods where we can engineer macro properties of a corpus, giving a range of models for each of the salient properties: document length distribution, word frequency distribution (for independent and non-independent cases), word length and textual representation, and corpus growth. We present results of emulating existing corpora and for scaling up corpora by two orders of magnitude. We show that simulated collections generated with relatively simple methods are suitable for some purposes and can be generated very quickly. Indeed it may sometimes be feasible to embed a simple lightweight corpus generator into an indexer for the purpose of efficiency studies. Naturally, a corpus of artificial text cannot support IR experimentation in the absence of a set of compatible queries. We discuss and experiment with published methods for query generation and query log emulation. We present a proof-of-the-pudding study in which we observe the predictive accuracy of efficiency and effectiveness results obtained on emulated versions of TREC corpora. The study includes three open-source retrieval systems and several TREC datasets. There is a trade-off between confidentiality and prediction accuracy and there are interesting interactions between retrieval systems and datasets. Our tentative conclusion is

that there are emulation methods which achieve useful prediction accuracy while providing a level of confidentiality adequate for many applications.

Modern society exists in a digital era in which high volumes of multimedia information exists. To optimize the management of this data, new methods are emerging for more efficient information retrieval. Web Semantics for Textual and Visual Information Retrieval is a pivotal reference source for the latest academic research on embedding and associating semantics with multimedia information to improve data retrieval techniques. Highlighting a range of pertinent topics such as automation, knowledge discovery, and social networking, this book is ideally designed for researchers, practitioners, students, and professionals interested in emerging trends in information retrieval.

Since user study design has been widely applied in search interactions and information retrieval (IR) systems evaluation studies, a deep reflection and meta-evaluation of interactive IR (IIR) user studies is critical for sharpening the instruments of IIR research and improving the reliability and validity of the conclusions drawn from IIR user studies. To this end, we developed a faceted framework for supporting user study design, reporting, and evaluation based on a systematic review of the state-of-the-art IIR research papers recently published in several top IR venues (n=462). Within the framework, we identify three major types of research focuses, extract and summarize facet values from specific cases, and highlight the under-reported user study components which may significantly affect the results of research. Then,

we employ the faceted framework in evaluating a series of IIR user studies against their respective research questions and explain the roles and impacts of the underlying connections and "collaborations" among different facet values. Through bridging diverse combinations of facet values with the study design decisions made for addressing research problems, the faceted framework can shed light on IIR user study design, reporting, and evaluation practices and help students and young researchers design and assess their own studies.

Multimodal Behavioral Analysis in the Wild: Advances and Challenges presents the state-of-the-art in behavioral signal processing using different data modalities, with a special focus on identifying the strengths and limitations of current technologies. The book focuses on audio and video modalities, while also emphasizing emerging modalities, such as accelerometer or proximity data. It covers tasks at different levels of complexity, from low level (speaker detection, sensorimotor links, source separation), through middle level (conversational group detection, addresser and addressee identification), and high level (personality and emotion recognition), providing insights on how to exploit inter-level and intra-level links. This is a valuable resource on the state-of-the- art and future research challenges of multi-modal behavioral analysis in the wild. It is suitable for researchers and graduate students in the fields of computer vision, audio processing, pattern recognition, machine learning and social signal processing. Gives a comprehensive collection of information on the state-of-the-art, limitations, and challenges associated with extracting behavioral cues from real-world scenarios Presents numerous applications on how different behavioral cues have been successfully extracted from different data sources Provides a wide variety of methodologies used to extract behavioral cues from multi-modal data

Personal Information Management (PIM) is the art of getting things done in our lives through information. How do we - can we better - manage our information at home, at school, at work, at play and "@large" in a global community? How do we use information not only to know but also to represent, communicate and effect useful change in the world around us? In the study of PIM, does the search for practical methods with practical impact lead to methods that are "massive open on-line"? Can the ancient practice of storytelling help us better to weave our fragmented information together? In the practice of PIM, how can our information best serve as "near knowledge" - close at hand and, through our information tools, serving in practical ways to extend the knowledge that's "in the head"? If attempts to multitask lead to ineffective, even dangerous, instances of task switching and divided attention, can better PIM help us to realize, instead, opportunities for "multi-goaling" where the same time and effort accomplishes not just one but several goals? These and other questions are addressed in this third and final book to conclude the series on "The Future of Personal Information Management". Part 1, "Our Information, Always and Forever", covered the fundamentals of PIM and then explored the seismic shift, already well underway, towards a world where our information is always at hand (thanks to our devices) and "forever" on the web. Part 2, "Transforming Technologies to Manage Our Information", provided a more focused look at technologies for managing information. The opening chapter discussed "natural interface" technologies of input/output to free us from keyboard, screen and mouse. Successive chapters then explored technologies to save, search and structure our information. A concluding chapter introduced the possibility that we may see dramatic reductions in the "clerical tax" we pay as we work with our information. Now in Part 3, "Building a Better World with Our Information", focus shifts to the practical

present and to the near future. Part 3 is in three chapters: • Group information management and the social fabric in PIM. How do we preserve and promote our PIM practices as we interact with others at home, at work, at play and in wider, even global, communities? (Chapter 10). • Designing for PIM in the development of tools and in the selection of teachable (learnable) "better practices" of PIM. (Chapter 11). • To each of us, our own concludes with an exploration of the ways each of us, individually, can develop better practices for the management of our information in service of the lives we wish to live and towards a better world we all must share. (Chapter 12).

Information Architecture is about organizing and simplifying information, designing and integrating information spaces/systems, and creating ways for people to find and interact with information content. Its goal is to help people understand and manage information and make the right decisions accordingly. This updated and revised edition of the book looks at integrated information spaces in the web context and beyond, with a focus on putting theories and principles into practice. In the ever-changing social, organizational, and technological contexts, information architects not only design individual information spaces (e.g., websites, software applications, and mobile devices), but also tackle strategic aggregation and integration of multiple information spaces across websites, channels, modalities, and platforms. Not only do they create predetermined navigation pathways, but they also provide tools and rules for people to organize information on their own and get connected with others. Information architects work with multi-disciplinary teams to determine the user experience strategy based on user needs and business goals, and make sure the strategy gets carried out by following the user-centered design (UCD) process via close collaboration with others.

Drawing on the authors' extensive experience as HCI researchers, User Experience Design practitioners, and Information Architecture instructors, this book provides a balanced view of the IA discipline by applying theories, design principles, and guidelines to IA and UX practices. It also covers advanced topics such as iterative design, UX decision support, and global and mobile IA considerations. Major revisions include moving away from a web-centric view toward multi-channel, multi-device experiences. Concepts such as responsive design, emerging design principles, and user-centered methods such as Agile, Lean UX, and Design Thinking are discussed and related to IA processes and practices.

The time-worn aphorism "close only counts in horseshoes and hand grenades" is clearly inadequate. Close also counts in golf, shuffleboard, archery, darts, curling, and other games of accuracy in which hitting the precise center of the target isn't to be expected every time, or in which we can expect to be driven from the target by skilled opponents. This book is not devoted to sports discussions, but to efficient algorithms for determining pairs of closely related web pages—and a few other situations in which we have found that inexact matching is good enough — where proximity suffices. We will not, however, attempt to be comprehensive in the investigation of probabilistic algorithms, approximation algorithms, or even techniques for organizing the discovery of nearest neighbors. We are more concerned with finding nearby neighbors; if they are not particularly close by, we are not particularly interested. In thinking of when approximation is sufficient, remember the oft-told joke about two campers sitting around after dinner. They hear noises coming towards them. One of them reaches for a pair of running shoes, and starts to don them. The second then notes that even with running shoes, they cannot hope to outrun a bear, to which the first notes that most likely the bear will be satiated

after catching the slower of them. We seek problems in which we don't need to be faster than the bear, just faster than the others fleeing the bear.

Information Retrieval (IR) models are a core component of IR research and IR systems. The past decade brought a consolidation of the family of IR models, which by 2000 consisted of relatively isolated views on TF-IDF (Term-Frequency times Inverse-Document-Frequency) as the weighting scheme in the vector-space model (VSM), the probabilistic relevance framework (PRF), the binary independence retrieval (BIR) model, BM25 (Best-Match Version 25, the main instantiation of the PRF/BIR), and language modelling (LM). Also, the early 2000s saw the arrival of divergence from randomness (DFR). Regarding intuition and simplicity, though LM is clear from a probabilistic point of view, several people stated: "It is easy to understand TF-IDF and BM25. For LM, however, we understand the math, but we do not fully understand why it works." This book takes a horizontal approach gathering the foundations of TF-IDF, PRF, BIR, Poisson, BM25, LM, probabilistic inference networks (PIN's), and divergence-based models. The aim is to create a consolidated and balanced view on the main models. A particular focus of this book is on the "relationships between models." This includes an overview over the main frameworks (PRF, logical IR, VSM, generalized VSM) and a pairing of TF-IDF with other models. It becomes evident that TF-IDF and LM measure the same, namely the dependence (overlap) between document and query. The Poisson probability helps to establish probabilistic, non-heuristic roots for TF-IDF, and the Poisson parameter, average term frequency, is a binding link between several retrieval models and model parameters. Table of Contents: List of Figures / Preface / Acknowledgments / Introduction / Foundations of IR Models / Relationships Between IR Models / Summary & Research Outlook / Bibliography /

Author's Biography / Index

Big data and human-computer information retrieval (HCIR) are changing IR. They capture the dynamic changes in the data and dynamic interactions of users with IR systems. A dynamic system is one which changes or adapts over time or a sequence of events. Many modern IR systems and data exhibit these characteristics which are largely ignored by conventional techniques. What is missing is an ability for the model to change over time and be responsive to stimulus. Documents, relevance, users and tasks all exhibit dynamic behavior that is captured in data sets typically collected over long time spans and models need to respond to these changes. Additionally, the size of modern datasets enforces limits on the amount of learning a system can achieve. Further to this, advances in IR interface, personalization and ad display demand models that can react to users in real time and in an intelligent, contextual way. In this book we provide a comprehensive and up-to-date introduction to Dynamic Information Retrieval Modeling, the statistical modeling of IR systems that can adapt to change. We define dynamics, what it means within the context of IR and highlight examples of problems where dynamics play an important role. We cover techniques ranging from classic relevance feedback to the latest applications of partially observable Markov decision processes (POMDPs) and a handful of useful algorithms and tools for solving IR problems incorporating dynamics. The theoretical component is based around the Markov Decision Process (MDP), a mathematical framework taken from the field of Artificial Intelligence (AI) that enables us to construct models that change according to sequential inputs. We define the framework and the algorithms commonly used to optimize over it and generalize it to the case where the inputs aren't reliable. We explore the topic of reinforcement learning more broadly and introduce

another tool known as a Multi-Armed Bandit which is useful for cases where exploring model parameters is beneficial. Following this we introduce theories and algorithms which can be used to incorporate dynamics into an IR model before presenting an array of state-of-the-art research that already does, such as in the areas of session search and online advertising. Change is at the heart of modern Information Retrieval systems and this book will help equip the reader with the tools and knowledge needed to understand Dynamic Information Retrieval Modeling.

Collaboration among scholars has always been recognized as a fundamental feature in scientific discovery. The ever-increasing diversity among disciplines and complexity of research problems make it more impelling to collaborate in order to keep up with the fast pace of innovation and advance knowledge. Along with the rapidly developing Internet communication technologies and the increasing popularity of social web, we have observed many important developments of scholarly collaboration on the academic social web. In this lecture, we review the rapid transformation of scholarly collaboration on various academic social web platforms, and examine how these platforms have facilitated academics throughout their research life cycle- from forming ideas, collecting data, authoring articles to disseminating findings. We refer to the term academic social web platforms in this lecture as a category of Web 2.0 tools or online platforms (such as CiteULike, Mendeley, academia.edu, and ResearchGate) that enable and facilitate scholarly information exchange and participation. We will also examine scholars' collaboration behaviors include sharing academic resources, exchanging opinions, following each other's research, keeping up with current research trends, and most importantly, building up their professional networks. Inspired by the model developed

by G. Olson, Olson, and Venolia (2000) on factors for successful scientific collaboration, our examination of the status of scholarly collaboration on academic social web has four emphases: technology readiness, coupling work, building common ground, and collaboration readiness. Finally, we will talk about the insights and challenges of all these online scholarly collaboration activities imposed to the research communities who are engaging in supporting online scholarly collaboration. This lecture aims to help researchers and practitioners to understand the development of scholarly collaboration on academic social web, and to build up an active community of scholars who are interested in this topic.

Information Retrieval performance measures are usually retrospective in nature, representing the effectiveness of an experimental process. However, in the sciences, phenomena may be predicted, given parameter values of the system. After developing a measure that can be applied retrospectively or can be predicted, performance of a system using a single term can be predicted given several different types of probabilistic distributions. Information Retrieval performance can be predicted with multiple terms, where statistical dependence between terms exists and is understood. These predictive models may be applied to realistic problems, and then the results may be used to validate the accuracy of the methods used. The application of metadata or index labels can be used to determine whether or not these features should be used in particular cases. Linguistic information, such as part-of-speech tag information, can increase the discrimination value of existing terminology and can be studied predictively. This work provides methods for measuring performance that may be used predictively. Means of predicting these performance measures are provided, both for the simple case of a single term in the query and for multiple terms. Methods of applying these

formulae are also suggested.

Society faces many challenges in workplaces, everyday life situations, and education contexts. Within information behavior research, there are often calls to bridge inclusiveness and for greater collaboration, with user-centered design approaches and, more specifically, participatory design practices. Collaboration and participation are essential in addressing contemporary societal challenges, designing creative information objects and processes, as well as developing spaces for learning, and information and research interventions. The intention is to improve access to information and the benefits to be gained from that. This also applies to bridging the digital divide and for embracing artificial intelligence. With regard to research and practices within information behavior, it is crucial to consider that all users should be involved. Many information activities (i.e., activities falling under the umbrella terms of information behavior and information practices) manifest through participation, and thus, methods such as participatory design may help unfold both information behavior and practices as well as the creation of information objects, new models, and theories. Information sharing is one of its core activities. For participatory design with its value set of democratic, inclusive, and open participation towards innovative practices in a diversity of contexts, it is essential to understand how information activities such as sharing

manifest itself. For information behavior studies it is essential to deepen understanding of how information sharing manifests in order to improve access to information and the use of information. Third Space is a physical, virtual, cognitive, and conceptual space where participants may negotiate, reflect, and form new knowledge and worldviews working toward creative, practical and applicable solutions, finding innovative, appropriate research methods, interpreting findings, proposing new theories, recommending next steps, and even designing solutions such as new information objects or services. Information sharing in participatory design manifests in tandem with many other information interaction activities and especially information and cognitive processing. Although there are practices of individual information sharing and information encountering, information sharing mostly relates to collaborative information behavior practices, creativity, and collective decision-making. Our purpose with this book is to enable students, researchers, and practitioners within a multi-disciplinary research field, including information studies and Human–Computer Interaction approaches, to gain a deeper understanding of how the core activity of information sharing in participatory design, in which Third Space may be a platform for information interaction, is taking place when using methods utilized in participatory design to address contemporary societal

challenges. This could also apply for information behavior studies using participatory design as methodology. We elaborate interpretations of core concepts such as participatory design, Third Space, information sharing, and collaborative information behavior, before discussing participatory design methods and processes in more depth. We also touch on information behavior, information practice, and other important concepts. Third Space, information sharing, and information interaction are discussed in some detail. A framework, with Third Space as a core intersecting zone, platform, and adaptive and creative space to study information sharing and other information behavior and interactions are suggested. As a tool to envision information behavior and suggest future practices, participatory design serves as a set of methods and tools in which new interpretations of the design of information behavior studies and eventually new information objects are being initiated involving multiple stakeholders in future information landscapes. For this purpose, we argue that Third Space can be used as an intersection zone to study information sharing and other information activities, but more importantly it can serve as a Third Space Information Behavior (TSIB) study framework where participatory design methodology and processes are applied to information behavior research studies and applications such as information objects, systems, and services with

recognition of the importance of situated awareness.

Citation analysis—the exploration of reference patterns in the scholarly and scientific literature—has long been applied in a number of social sciences to study research impact, knowledge flows, and knowledge networks. It has important information science applications as well, particularly in knowledge representation and in information retrieval. Recent years have seen a burgeoning interest in citation analysis to help address research, management, or information service issues such as university rankings, research evaluation, or knowledge domain visualization. This renewed and growing interest stems from significant improvements in the availability and accessibility of digital bibliographic data (both citation and full text) and of relevant computer technologies. The former provides large amounts of data and the latter the necessary tools for researchers to conduct new types of large-scale citation analysis, even without special access to special data collections. Exciting new developments are emerging this way in many aspects of citation analysis. This book critically examines both theory and practical techniques of citation network analysis and visualization, one of the two main types of citation analysis (the other being evaluative citation analysis). To set the context for its main theme, the book begins with a discussion of the foundations of citation analysis in general, including an overview of what can and

what cannot be done with citation analysis (Chapter 1). An in-depth examination of the generally accepted steps and procedures for citation network analysis follows, including the concepts and techniques that are associated with each step (Chapter 2). Individual issues that are particularly important in citation network analysis are then scrutinized, namely: field delineation and data sources for citation analysis (Chapter 3); disambiguation of names and references (Chapter 4); and visualization of citation networks (Chapter 5). Sufficient technical detail is provided in each chapter so the book can serve as a practical how-to guide to conducting citation network analysis and visualization studies. While the discussion of most of the topics in this book applies to all types of citation analysis, the structure of the text and the details of procedures, examples, and tools covered here are geared to citation network analysis rather than evaluative citation analysis. This conscious choice was based on the authors' observation that, compared to evaluative citation analysis, citation network analysis has not been covered nearly as well by dedicated books, despite the fact that it has not been subject to nearly as much severe criticism and has been substantially enriched in recent years with new theory and techniques from research areas such as network science, social network analysis, or information visualization. Visual Information Retrieval using Java and LIREMorgan & Claypool Publishers

A trustworthy repository provides assurance in the form of management documents, event logs, and audit trails that digital objects are being managed correctly. The assurance includes plans for the sustainability of the repository, the accession of digital records, the management of technology evolution, and the mitigation of the risk of data loss. A detailed assessment is provided by the ISO-16363:2012 standard, "Space data and information transfer systems—Audit and certification of trustworthy digital repositories." This book examines whether the ISO specification for trustworthiness can be enforced by computer actionable policies. An implementation of the policies is provided and the policies are sorted into categories for procedures to manage externally generated documents, specify repository parameters, specify preservation metadata attributes, specify audit mechanisms for all preservation actions, specify control of preservation operations, and control preservation properties as technology evolves. An application of the resulting procedures is made to enforce trustworthiness within National Science Foundation data management plans.

In this book, we aim to provide a fairly comprehensive overview of the scalability and efficiency challenges in large-scale web search engines. More specifically, we cover the issues involved in the design of three separate systems that are commonly available in every web-scale search engine: web crawling, indexing,

and query processing systems. We present the performance challenges encountered in these systems and review a wide range of design alternatives employed as solution to these challenges, specifically focusing on algorithmic and architectural optimizations. We discuss the available optimizations at different computational granularities, ranging from a single computer node to a collection of data centers. We provide some hints to both the practitioners and theoreticians involved in the field about the way large-scale web search engines operate and the adopted design choices. Moreover, we survey the efficiency literature, providing pointers to a large number of relatively important research papers. Finally, we discuss some open research problems in the context of search engine efficiency.

Copyright: 551a34344212f627b3939bdc2fff87cf