Aug 10th, 12:00 AM

The Use of Text Analytics to Investigate Concepts in Intra- and Inter-disciplinary Software Piracy Research

Jerald Hughes
University of Texas, Rio Grande Valley, j.hughes@utrgv.edu

Juan A. Chavarria
Emporia State University, jchavar1@emporia.edu

Francis Kofi Andoh-Baidoo
University of Texas Rio Grande Valley, francis.andohbaidoo@utrgv.edu

Use of Text Analytics for Shared Concepts Identification in Intra and Interdisciplinary Research: The Case of Software Piracy

Completed Research

Jerald K. Hughes
University of Texas Rio Grande Valley
j.hughes@utrgv.edu

Juan A. Chavarria
Emporia State University
jchavar1@emporia.edu

Francis Kofi Andoh-Baidoo
University of Texas Rio Grande Valley
francis.andohbaidoo@utrgv.edu

Abstract

IS research has linked collaborators from diverse domains. IS research requires selecting and addressing an appropriate intradisciplinary or interdisciplinary scope. Identifying gaps in the current literature and deciding when and how collaborations among different disciplines may be fruitful poses challenges. We propose a process to analyze a corpus of documents from any topic, to identify potential collaboration areas.

A text analytics process is used to find areas of commonality and exclusivity among questions addressed in existing IS work by analyzing abstracts in papers from multiple disciplines studying 'software piracy.' We use term-term co-occurrence to find all the terms used in close proximity to the topic. We identify which terms are most prominent in each discipline, show quantitatively how these usages coincide or diverge across disciplines, measure the overlap between pairs of disciplines, and identify clusters of terms shared among disciplines. Specific findings from this case of software piracy are presented.

Keywords

Text Analytics, Interdisciplinary, Knowledge Contribution, Software Piracy.

Introduction

Senior Information Systems (IS) scholars agree that the nature of the discipline dictates the need for collaboration to build cumulative knowledge and to enhance the field's legitimacy (e.g., Agarwal and Dhar 2014; Nunamaker et al. 2017). The use of analytics in several disciplines on diverse phenomena presents new opportunities for the IS field to exploit the benefits of cross- and transdisciplinary knowledge contribution on various topics (Agarwal and Dhar 2014). In view of this, researchers have examined literature to present potential future research agendas on various topics including software piracy (Chavarria et al. 2016), and Privacy (Bélanger and Crossler 2011). While such studies have typically relied on manual content analysis, analytics tools open additional possibilities for studying cross- and transdisciplinary knowledge contributions. We demonstrate the usefulness of a text analytics approach, applied here on a single specific topic chosen as an example context.

Data mining techniques for text analytics enable researchers to obtain enhanced insights into how terms in the abstracts relate to each other, including both how term usages overlap among multiple disciplines, and how they are used exclusively within each discipline, especially when studying intradisciplinary and inter-disciplinary topic areas. We use custom software to measure co-occurrences of salient terms of a topic of interest. The relative frequencies of terms in each study which are matched in various papers with the central topic will show how studies from different disciplines either intersect with other disciplines or deploy exclusive usages within each discipline individually. The objectives of this paper are thus to investigate how: (1) papers of various disciplines overlap and non-overlap in their usage of the "central topic" term and its variants, and (2) researchers can target specific related concept clusters to develop inter-

subjective and fused concepts for potential interdisciplinary knowledge contributions in a chosen research area.

While intradisciplinary research employs concepts of the IS discipline itself, an interdisciplinary knowledge contribution in this context is characterized as "An integrated contribution that develops fundamentally new, fused concepts, which exist at the intersection and inter-subjective areas of IS and other disciplines, and that draws from (i) the IS discipline, and (ii) other disciplines" (Tarafdar and Davison 2018, p. 527). Empirical studies that survey the nature of intradisciplinary and interdisciplinary knowledge contribution include those that have examined the entire IS field (e.g., Tarafdar and Davison 2018), or specific IS sub-disciplines such as software piracy (Chavarria et al. 2016), economics of IT (Brynjolfsson 2021), and Digitalization (Parmiggiani et al. 2020). These studies found that products of the majority of IS research have been intradisciplinary, with fewer interdisciplinary contributions (Chavarria et al. 2016; Oh et al. 2005; Tarafdar and Davison 2018). These results have led to a call for more interdisciplinary knowledge contributions in the IS field (Tarafdar and Davison 2018) and its sub-disciplines (Chavarria et al. 2016).

Understanding where IS researchers have already made interdisciplinary contributions will help in increasing interdisciplinary research. Yet less is known about which other topics currently addressed in IS research may also have potential for development toward interdisciplinary knowledge contributions, if such topics are also separately receiving attention in other disciplines. To identify such areas of potential, it would be useful to examine multiple disciplines, searching for areas of overlap with IS, as well as those currently unique to IS. Having such understanding will help IS scholars to collaborate with others in related disciplines on the fused concepts that can lead to developing novel theories that explain phenomena of mutual benefit to IS and other disciplines.

To demonstrate the usefulness of our approach, we use the context of software piracy as a central topic present in many disciplines (Chavarria et al. 2016). We search for potential areas of overlap across disciplines by using text analytics tools to investigate the topic of software piracy as a case example of our approach. A recent literature review showed that software piracy is a cross-disciplinary research area (Chavarria et al. 2016) (with 'cross-disciplinary' in this context taking the dictionary meaning of relating to more than one discipline). However, we do not know the degree to which the different disciplines associated with software piracy research may overlap in terms of the concepts employed. Some concepts relevant to software piracy may be closely bound to a single discipline, while others have broader applicability in multiple disciplines. This is the focus of the current work. Thus, to investigate the main research foci of software piracy research, we employ techniques for automated text analytics on the abstracts of 162 journal articles from 11 disciplines.

Our study contributes to both research and practice. In terms of research, first we offer an approach for studying the intersecting and overlapping terminologies, to identify areas providing opportunities for interdisciplinary knowledge contributions. The text analytics approach proposed here can be used to study interdisciplinary knowledge contributions in other sub-domains, or the entire IS field and beyond. Second, we expand our knowledge on software piracy research itself by presenting new insights into both the specific terms employed in current areas of intradisciplinary work, and those with potential for interdisciplinary software piracy research. For practice, text analytics software as demonstrated here can be useful for abstracting actionable knowledge, which is embodied in current published research, particularly when that knowledge is distributed across different disciplines. These tools and techniques can be expanded to build multiple sets of relevant topic terms for software piracy research and other collaborative communities. Then, funding organizations and software piracy industry practitioners can identify which collaborative teams may be appropriate to address specific issues based on these sets of terms, depending on their interests and problems.

The rest of the paper is organized as follows. In the next section, we present relevant research on intradisciplinary and interdisciplinary knowledge contributions. We then describe the methodology that employs automated text analytics to classify software piracy knowledge contributions. We present our findings and then conclude with limitations and potential future research areas.

Relevant Background

In this paper we employ the theoretical framework of intradisciplinary and interdisciplinary knowledge contribution to encompass multiple fields of research, where IS researchers collaborate with researchers

from other disciplines to generate scientific knowledge (Tarafdar and Davison 2018). We look first at the issues of collaboration among researchers.

Collaborative Research

In academic research there is increased interest in collaborative teams (Jiang et al. 2018). These have several benefits. The following are identified as drivers of high level of research collaboration: "the rapid increase in the amount of information available, the wide variety of new research techniques, the myriad sources of interdisciplinary funding, the high levels of diverse statistical skills and computation resources required, the huge amount of specific in-depth knowledge of subject areas and proprietary programs necessary to compete in large scientific enterprises, and of course the emergence of large interdisciplinary institutes and teams" (McDermott and Hatemi 2010, p. 49). In such projects, the question arises of how different members are to coordinate their ideas. Research on team collaboration suggests that the knowledge of shared concepts is useful in ensuring team success (Roschelle and Teasley 1995). Shared understanding and common goals among actors facilitate the creation of intellectual capital and knowledge sharing in a network (Chiu et al. 2006; Tsai and Ghoshal 1998; Wasko and Faraj 2005). In all research environments, including those involving only a single discipline, shared concepts are critical: "It is well established that successful biomedical, uni-disciplinary research benefits from a clear set of shared concepts." (O'Sullivan et al. 2010, p. 1175). Other researchers state that "shared concepts provide focus over time and foster discussions among participants and across communities" (Baker et al. 2005, p. 6). Research collaborations with team members of a single discipline already have a strong basis for shared concepts characteristic of their discipline. But collaborations across multiple disciplines to create interdisciplinary research products raises the question of how researchers in disparate areas can carry out processes of discovery and development for shared conceptualizations. We turn now to these issues of intradisciplinary and interdisciplinary research.

Intra and Interdisciplinary IS research

Scholars acknowledge the importance of collaborative research in IS discipline legitimacy (e.g., Agarwal and Dhar 2014; Nunamaker et al. 2017). To support such collaborative work, Tarafdar and Davison (2018) developed a theoretical framework to conceptualize IS knowledge contributions. The framework classified IS knowledge contributions into distinct categories of intradisciplinary and interdisciplinary. Intradisciplinary was further divided into single disciplinary and home disciplinary, while interdisciplinary was split into subcategories of cross disciplinary and interdisciplinary. While single disciplinary and home disciplinary categories provide knowledge contributions to the IS discipline, single disciplinary knowledge contributions are solely based on core IS concepts and theories, while home disciplinary contributions draw concepts and theories from both IS and other reference disciplines. Cross disciplinary research generates knowledge for IS and for other disciplines by drawing on theories and concepts from IS and other reference disciplines. Finally, interdisciplinary research leads to an integrated knowledge contribution by drawing from IS and other disciplines. This last type is the most integrated in terms of the ideas employed: "fused concepts existing at the intersection and inter-subjective areas of IS and other disciplines" (Tarafdar and Davison 2018, p. 528). The authors have called specifically for more interdisciplinary research, while also providing guidelines for producing all four types of knowledge contributions. Recent work has demonstrated the usefulness of these concepts, and the need for interdisciplinary research, e.g., in marginalization (Chughtai et al. 2020), sustainable tourism (Nunkoo et al. 2021), and technostress (Ali et al. 2021). To further investigate these issues in a case of examining a specific domain of interest, we focus now on a topic which, while prominent in IS research, has broad representation in many disciplines, software piracy.

Software Piracy—Research

Most reviews of software piracy research have focused on a specific viewpoint, such as criminal (Holsapple et al. 2008) or ethical (Siponen and Vartiainen 2004), or investigated a specific category of potential pirates (Liang and Yan 2005). However, according to Chavarria et al. (2016), these studies did not profile the accumulated literature on software piracy across different disciplines. Software piracy emerges as a useful case for our application of text analytics across multiple disciplines because of the particularly broad array of disciplines publishing research addressing software piracy as a central topic. Teston (2008), in a study of software piracy among students, points out, for example, that concepts of digital property are interdisciplinary; for his study, interdisciplinarity presents multiple opportunities for pedagogy integrated

across the curriculum. Willison and Siponen (2008) specifically called out the usefulness of concepts from a discipline outside Information Systems – Criminology – in order to understand and prevent software piracy. We expand on this by looking at many disciplines which have studied this central topic. Chavarria et al. (2016) presents a comprehensive literature review demonstrating that software piracy research includes research from IS and other disciplines: economics, marketing, management, psychology, sociology, communications, criminology, education, engineering, and law. The authors identify the top three disciplines that study software piracy as information systems, management, and economics, with contributions of 37.43%, 27.93%, 10.06% respectively, with these three constituting 75.42% of the total contributions.

In this study, we show how the diverse disciplines overlap in the terminologies related to software piracy that interest these groups, by examining how these terms are expressed in the published abstracts. We also determine which terminologies are unique to individual disciplines – terms not shared across disciplines. Thus, we demonstrate, using text analytics, the extent of intradisciplinary and interdisciplinary knowledge contributions of software piracy research. Such knowledge can help build a better understanding of the different combinations of disciplines that study software piracy. The following section describes the processes employed.

Methodology

Our research method involves preliminary work to clean and prepare the texts, followed by three analysis steps: initial measures (stage 1), terms analysis (stage 2) and clusters analysis (stage 3). First is the preparation. Papers of the 'Software Piracy' publication from Chavarria et al. (2016) et al. constitute a corpus of research on software piracy, drawn from multiple disciplines. The texts for this analysis are abstracts, which provide the most direct descriptions of the treatment of the topic 'software piracy', relevant to our research objectives (Ou et al. 2008). Non-standard characters are replaced with equivalents (e.g., image of 'A' replaced with character 'A'), then the full texts are re-encoded in Unicode UTF-8 and stored individually as .txt files.

Next is the measures phase (stage 1 analytics). All .txt files are scanned for terms used at multiple levels of immediate proximity with 'piracy'. Counts of co-occurring terms are collected individually for distances from the target term of 1,2,3,4 and 5. The number of times each pair appears in a document is stored. The categories of disciplines for the source documents already established in Chavarria et al. (2016) are used to categorize each document as a member of a discipline. This category variable is included in order to enable later analyses which identify common terms among disciplines, which could then indicate a conceptual basis for collaborative research. One row from this stage of analysis, for example, yields 6 uses of 'enforcement' within 5 words before or after the term 'piracy', in an Information Systems paper (August and Tunca 2008). The completed output from this stage provides similar rows for every word used within a distance of 5 words of the target term 'piracy'.

Next is the terms analysis phase (stage 2 analytics). Tabulation of pairs is repeated for variants, including for example 'pirate', 'pirating', 'pirates', etc. A term used as synonymous to the core term was discovered, 'softlifting', so counts of matched pairs with this term were also measured. The counts of usages for all documents together are calculated. Term-pairs characteristic of each discipline are identified. Counts are summed for each of the term-pairs, across all documents in that discipline. This yields 11 lists of term pairs by discipline: Information Systems (Info), Management (Mgmt), Economics (Econ), Marketing (Mark), Psychology (Psych), Education (Educ), Criminology (Crim), Law, Sociology (Soci), Engineering, and Communication. The lowest document count disciplines of Engineering and Communication (2 each) are not included in the remaining steps. Tables representing the usages of these terms within and among disciplines are created. Data visualizations in the form of word clouds are produced for human inspection of prominent elements among the disciplines.

Finally, in the clusters phase (stage 3 analytics), results are inspected for clusters of disciplines which share terms closely associated with 'piracy' or 'softlifting'. This analysis allows us to specify groups of disciplines (e.g., Info, Mgmt, Econ) which have important terms in common, indicating shared specific interests in the domain of software piracy. In cases where researchers have themselves identified topics of potential interdisciplinary potential, but where terms relevant to those topics are not yet present in the shared lists, a fruitful gap in the literature may be present. We present two examples of such clusters, including both shared terms 2-way and 3-way, and terms exclusive to each member of the cluster.

Results and Discussion

Table 1 presents the terms associated with 'piracy' and its variants at a proximity <= 5 before or after, ranked according to how many disciplines an associated term appeared in (3rd column). For example, it is no surprise that 'research' appears in every discipline represented, since the corpus is academic journals. Also near the top ranks are 'implication', 'finding' and 'study', words commonly used in research abstracts. However, notice that at the top of the ranked list, we see specific topics and themes associated with 'piracy' in many studies, such as 'computer', 'data', 'attitudes', 'industry', 'moral', and 'social', appear in many disciplines. These are potential themes for broadly interdisciplinary research. Further down, terms like 'economic', 'property', and 'legal', while present in five disciplines, are missing from the other disciplines in this corpus. The complete list, while far too long to present here, provides information about which terms are common to only two or a few disciplines, such as 'corruption' (Mark, Mgmt), 'productivity' (Info, Econ), or 'gender' (Crim, Educ, Psych). Finally, the largest portion of the list, 948 rows, consists of terms which appeared only in one discipline, such as 'entertainment' (Educ), 'collectivist' (Econ), or 'felony' (Crim). At its top ranks, this table points toward the specific aspects of software piracy which appear most widely across all disciplines present; researchers considering interdisciplinary research with the broadest implications, participation, and potential impacts may wish to address these aspects. In the middle ranks of Table 1, researchers who have already identified a particular angle which they wish to study within the central topic of software piracy can find in the discipline columns potential related theories, findings, and thus possible collaborators from other fields.

#	Rank	Discipline Count	Related Term	Disc1	Disc2	Disc3	Disc4	Disc5	Disc6	Disc7	Disc8	Disc9
1	1	9	research	Crim	Econ	Educ	Info	Law	Mark	Mgmt	Psych	Soci
2	1	9	software	Crim	Econ	Educ	Info	Law	Mark	Mgmt	Psych	Soci
3	2	8	computer	Crim	Econ	Educ	Info	Mark	Mgmt	Psych	Soci	
4	2	8	using	Crim	Econ	Educ	Info	Mark	Mgmt	Psych	Soci	
5	3	7	data	Crim	Econ	Info	Mark	Mgmt	Psych	Soci		
6	3	7	findings	Crim	Econ	Educ	Info	Mgmt	Psych	Soci		
7	3	7	study	Crim	Econ	Educ	Info	Mark	Mgmt	Psych		
8	4	6	affect	Econ	Educ	Info	Mark	Mgmt	Psych			
9	4	6	attitudes	Crim	Educ	Info	Mgmt	Psych	Soci			
10	4	6	factors	Crim	Educ	Info	Mark	Mgmt	Psych			
11	4	6	impact	Econ	Educ	Info	Law	Mark	Mgmt			
12	4	6	implications	Crim	Econ	Educ	Info	Mark	Mgmt			
13	4	6	Important	Crim	Econ	Educ	Info	Mark	Mgmt			
14	4	6	industry	Econ	Info	Law	Mark	Mgmt	Psych			
15	4	6	level	Econ	Info	Law	Mark	Mgmt	Soci			
16	4	6	low	Crim	Econ	Info	Mgmt	Psych	Soci			
17	4	6	measures	Crim	Econ	Info	Law	Mgmt	Psych			
18	4	6	moral	Educ	Info	Mark	Mgmt	Psych	Soci			
19	4	6	paper	Econ	Educ	Info	Law	Mark	Mgmt			
20	4	6	rates	Econ	Info	Law	Mark	Mgmt	Soci			
21	4	6	related	Econ	Educ	Info	Law	Mgmt	Psych			
22	4	6	social	Crim	Econ	Info	Mark	Mgmt	Soci			
23	4	6	students	Crim	Econ	Educ	Info	Mgmt	Psych			
24	5	5	analysis	Crim	Econ	Info	Mark	Mgmt				
25	5	5	attitude	Educ	Info	Mark	Mgmt	Psych				
26	5	5	behavior	Educ	Info	Law	Mark	Mgmt				
27	5	5	business	Educ	Info	Law	Mark	Mgmt				
28	5	5	control	Crim	Info	Mgmt	Psych	Soci				
29	5	5	cost	Econ	Info	Law	Mgmt	Psych				
30	5	5	countries	Econ	Info	Law	Mgmt	Soci				
31	5	5	decision	Econ	Educ	Info	Mgmt	Psych				
32	5	5	due	Econ	Info	Law	Mark	Mgmt				
33	5	5	economic	Econ	Info	Law	Mark	Mgmt				
34	5	5	effect	Crim	Econ	Info	Law	Mgmt				
35	5	5	empirical	Crim	Econ	Educ	Info	Mgmt				
36	5	5	examines	Econ	Law	Mark	Mgmt	Psych				

37	5	5	legal	Econ	Educ	Info	Mark	Mgmt				
38	5	5	literature	Econ	Info	Mark	Mgmt	Psych				
39	5	5	lower	Econ	Info	Law	Mgmt	Soci				
40	5	5	national	Econ	Info	Law	Mark	Mgmt				
41	5	5	new	Info	Law	Mark	Mgmt	Psych				
42	5	5	property	Econ	Info	Mark	Mgmt	Soci				
43	5	5	rate	Econ	Info	Law	Mark	Mgmt				
44	5	5	role	Crim	Econ	Info	Mgmt	Psych				
45	5	5	sample	Crim	Econ	Educ	Mgmt	Soci				
46	5	5	self	Crim	Info	Mgmt	Psych	Soci				
47	5	5	survey	Educ	Info	Law	Mark	Mgmt				
48	5	5	theory	Crim	Info	Mark	Mgmt	Psych				
49	5	5	use	Educ	Info	Mgmt	Psych	Soci				

Table 1. Terms Associated With 'Piracy' Across Disciplines and Its Variants At A Proximity <=5

Someone researching software piracy across different 'countries', for example, can see from rank 30 that Econ, Law, Management and Sociology all have contributions. At the ranks towards the bottom of the full listing (not presented here) one may find specific terms showing current boundaries of practice of those researchers investigating software piracy. Conversely, researchers starting out from the premise of addressing potential intersections with a particular separate discipline can find in this table specific terms which identify those crossover points.

Table 2 lists the top 10 terms co-occurring with 'piracy' or its variants, with their total counts for each of the top four major disciplines. Several terms in Information Systems are typical behavioral research issues, but we also see social, global, moral and economic in the top 10. For management, the focus is on benefits, government, price, governance and policy issues. For economics, issues pertinent to modeling buyers and sellers are captured, but notice also that 'deterring' is included here, potentially relevant to the goals of other researchers.

INFO		MGMT		ECON		MARK	
rate	84	Benefits	10	Monopolist	4	purchasing	4
attitude	53	Norm	9	Responsible	3	managers	2
intention	49	Diffusion	7	Tolerate	3	america	2
behavior	48	government	7	Commercial	3	functionality	2
level	37	Price	6	Liability	3	latin	2
social	28	unauthorized	6	Buyer	2	multinational	2
consumer	26	corruption	5	Deterring	2	tolerating	2
global	22	governance	5	Economy	2	consciousness	1
moral	20	Adopter	4	Incentive	2	invisible	1
economic	20	Leniency	4	Manufacturer	2	liberalization	1

Table 2. Top 10 Usage Terms by Discipline

At the same time, specific attitudes towards piracy span multiple disciplines: 'moral' in INFO, 'unauthorized', 'corruption' and 'leniency' in MGMT, 'responsible and 'tolerate' in ECON, and 'tolerating' and 'liberalization' in MARK. We identify these items as a cluster of related terms with potential for development of fused concepts across disciplines, per our second research question. Another example is the intersection point visible with 'global' and 'multinational' linking IS and Marketing.

The word clouds, in Figures 2, 3 and 4, provide an overview of terms appearing prominently in each discipline's subset of the literature. For example, notice that the Information Systems cloud includes 'moral' and the phrase 'moral intensity', and that the Criminal Justice cloud also includes 'moral'. Information Systems shares a different set of terms with the Psychology cloud: both include 'gender'; Psychology also uses 'moral intensity', indicating interest in the moral dimensions among all three disciplines. These and other shared terms become potential candidates for areas which might prove fruitful for the development of fused concepts residing at the intersection of IS with other disciplines.

Figure 2. Information Systems Discipline Word Cloud

Figure 3. Criminal Justice Discipline Word Cloud

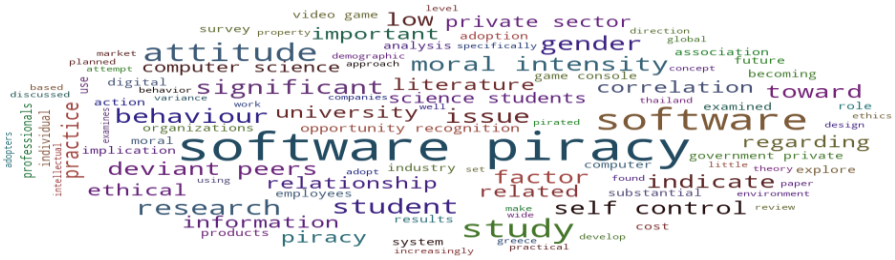
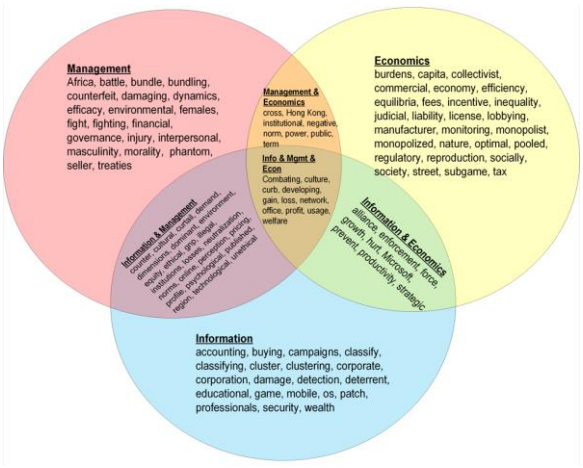


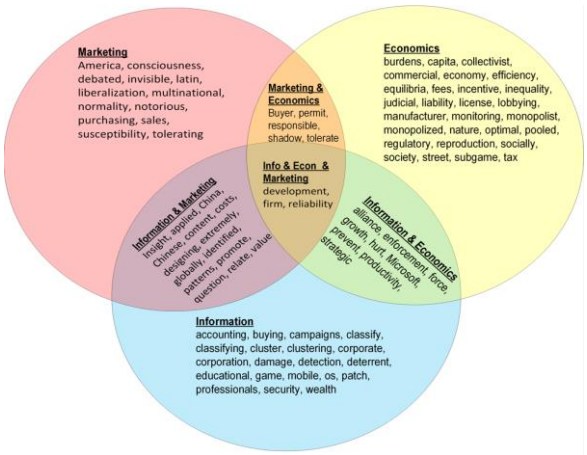
Figure 4. Psychology Discipline Word Cloud

Once we have identified the terms that overlap and do not overlap, per our first research question, we can begin to build clusters of terms which are distinct for each discipline, and for each intersection of that discipline with others. The Venn diagrams in Figure 5 shows such distinctions. These figures demonstrate how our approach can facilitate the development of term clusters for this topic domain.

The diagram on the left of Figure 5 presents the exclusively used terms for each of the disciplines in the outer sectors of the Venn diagram, addressing the non-overlap portion of our first research question. Thus, for example in this corpus only Information Systems papers used terms such as ‘game’, ‘mobile’, and ‘os’ with “piracy” and its variants. For relatively new and swiftly developing technology areas such as ‘mobile’, this particular term might signal a useful gap in literature, for studies which aim to be interdisciplinary. For topic modeling in text analytics, the exclusive terms would be candidates for a text analytics dictionary linking bodies of text as belonging to that topic. Notice that this set of terms is specific to ‘piracy’, which is a more detailed view of what the Information Systems discipline alone provides. Both Information Systems and Management papers, by contrast, discuss ethical issues: the terms ‘ethical’, ‘unethical’, ‘norm’, and ‘neutralization’ appear in both disciplines. This overlap is revealed by the analysis technique to be unique to these two disciplines. This could indicate an established area within software piracy research but could also alert researchers to possible gaps of import for new studies: perhaps Law, Psychology, or Criminology could have new contributions to make regarding software piracy’s ethical aspects. The center of the left diagram on Figure 5 presents the 3-way intersection of Information, Management and Economics, showing the overlap of associations with software piracy for: ‘profit’, ‘gain’, and ‘loss’, but also ‘combating’ and ‘culture’, indicating how the specific economic outcomes related to software piracy impact stakeholders in firms, as well as the wider public. These connections already exist in this corpus of software piracy literature, thereby demonstrating quantitatively a dimension of important interdisciplinary work on piracy.



Info, Mgmt and Econ



Info, Mark and Econ

Figure 5. Exclusive and Exclusively Shared Terms

On the right-side diagram of Figure 5, however, notice that the 3-way intersection list has fewer terms indicating a shared topic—if there even is one. The term 'reliability', likely appears only as a metric of the methodology, leaving only 'development' and 'firm' as possible indicators. This might show that a gap exists in interdisciplinary investigation involving Economics, Marketing, and Information Systems. From these sectors in both figures, we can see that the corpus addresses Africa only in the Management set, Hong Kong appears only in the intersection of Mgmt and Econ, and China appears only in the intersection of Info and Mark. Thus, a researcher wishing to address international aspects of software piracy in critical locations around the globe may wish to explore these.

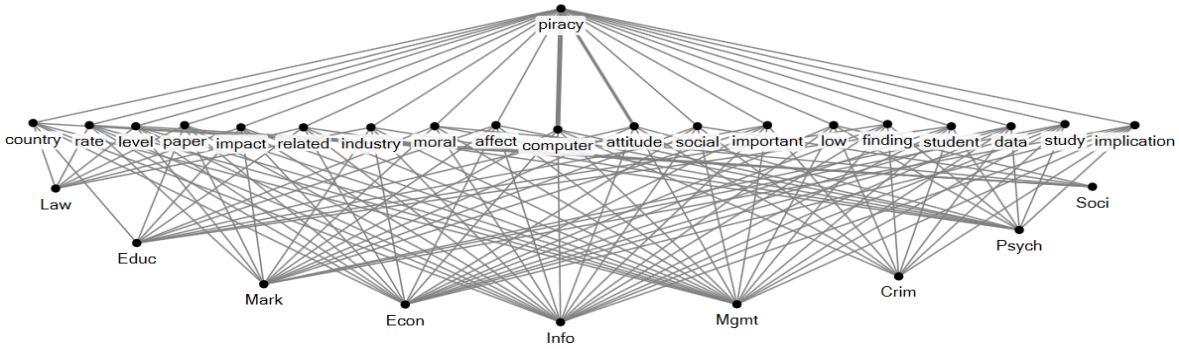


Figure 6. Software Piracy Research Terms Network Diagram

Towards a future analysis of measures like the ones presented in this paper, Figure 6 illustrates a network graph of the most used terms across all disciplines. Visualizations and network graph analytics techniques can reveal relationships such as those visible here, between 'piracy' (the central topic overall), terms associated with the central topic (e.g., moral, social, attitude), and the disciplines which address them (Info, Econ, etc.). Here we immediately see the differences between disciplines, for the most frequent terms associated with 'piracy'; Info and Mgmt are connected to all the top terms, whereas Soci and Law have fewer connections. Information Systems, for 'software piracy', constitutes a primary central nexus in itself, a discipline around which the ideas and terms of the other disciplines listed cluster.

Conclusions and Implications

In this study, we employ text analytics to drill down into the specific text usages of an interdisciplinary domain of research, software piracy. By providing details of the textual content of the abstracts of the studies, we have identified the extent of overlap and non-overlap between the sets of text contents in the various disciplines involved. In addition, we have delved into the specific contents (individual terms associated with 'piracy') which constitute the overlaps and non-overlaps. Our methodology presents tools for automated text analytics, which we believe have enormous potential for generating new insights through quantitative measures for text analysis.

This study conducts empirical quantitative exploratory analysis of the abstracts of 162 journal articles from 11 disciplines, on software piracy. Our results show how studies from different disciplines both intersect with other disciplines and generate exclusive usages within each discipline individually. Our findings reveal areas of common and exclusive interests in software piracy for this document corpus. The generation of the specific shared and unique concepts can lead to the development of different dictionaries for the diverse collaborative groups of software piracy researchers. We offer implications for research and practice.

Implications for theory and Practice

The study that we present in this paper offers several implications to both research and practice. In terms of theory, first we present an analysis which quantitatively unpacks the concrete meaning and content of the topic of software piracy in academic articles from multiple disciplines. While significant development has already taken place in text analytics measuring occurrence of terms in documents, in this context our study is the first to develop a highly granular analysis which investigates the text corpus via term-term relationships. Also in terms of theory, we identify multiple sets of candidate terms for the creation of text analytics dictionaries on 'software piracy' relative to researchers' disciplines, as well as those which may constitute common understandings of the phenomena in question. Examples of takeaways relevant specifically to 'software piracy' include both positive findings: e.g., studies addressing the 'moral' aspects of software piracy are central to at least six disciplines, indicating broad interest and importance; and negative findings: e.g., 'mobile' appeared in this analysis exclusively in Information Systems journals, but should have strong potential for intersections with other disciplines, and thus constitutes a possible gap of interest. Follow-up work of the kind deploying text analytics such as those we use here could collect new data and perform analyses to shed light longitudinally on how interdisciplinary work may begin within one discipline, and then spread to others.

The techniques used here can be deployed and extended to any set of disciplines with potential for interdisciplinary collaborations. A particular benefit of employing automated text analytics for these purposes is the extraordinary scalability of the tools used. Analyses extending to many hundreds, to thousands, or even tens of thousands of documents—studies of a scope which put them out of reach for traditional eyes-on qualitative techniques which could require man-years of reading and re-reading just for a single study—become possible with the help of text analytics software. Finally, this type of analysis can be applied to help build communities of research within institutions such as universities. For example, a team seeking to be awarded a grant may need multiple domains of discipline expertise. The analysis method used here highlights the potential interdisciplinary domains which might facilitate the creation of the most effective teams by discipline and expertise. Organizations and industry practitioners can also use these techniques to develop deeper understandings of how critical concepts such as 'software piracy' are employed within their firms. Critically, an analysis of how staff dealing with software and possible legal implications discuss and communicate around issues of piracy may be able to flag problematic attitudes and practices at an early stage. Finally, funding organizations interested in supporting interdisciplinary research teams can benefit from this study in terms of what concepts constitute the intersection of interest and understanding.

Limitations and Future Research

One limitation of the findings is that this study used abstracts of research papers as the text corpus analyzed. We agree that the entire research article itself reflects the most complete record of the work that authors present on their chosen topics. However, we believe since researchers are restricted in terms of the number of words that they use in the abstract, they are necessarily careful in choosing which important words and issues to present in the abstract. In addition, a typical abstract includes the motivation for the study, the method used, the findings, the implications and limitations. Hence the authors typically address all the components of a research paper but in a more summarized form in the abstract. Future research could extend the tools and methods presented here, to use entire research articles for comparative analysis purposes. Further important future research could develop detailed dictionaries for terminologies for each unique discipline and the intersection of the various disciplines that are present in the software piracy research community, as well as for other topics.

References

Agarwal, R., and Dhar, V. 2014. "Big Data, Data Science, and Analytics: The Opportunity and Challenge for Is Research." INFORMS.

Ali, A., Jan, S., and Iqbal, P.M. 2021. "A Bibliometric Analysis on Technostress: An Emerging Discipline," *Journal of Indian Library Association* (57:4), pp. 37-48.

August, T., and Tunca, T.I. 2008. "Let the Pirates Patch? An Economic Analysis of Software Security Patch Restrictions," *Information Systems Research* (19:1), pp. 48-70.

Baker, K.S., Jackson, S.J., and Wanetick, J.R. 2005. "Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment," *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*: IEEE, pp. 219b-219b.

Bélanger, F., and Crossler, R.E. 2011. "Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems," *MIS Quarterly* (35:4), pp. 1017-1042.

Chavarria, J.A., Andoh-Baidoo, F.K., Midha, V., and Hughes, J. 2016. "Software Piracy Research: A Cross-Disciplinary Systematic Review," *Communications of the Association for Information Systems* (38).

Chiu, C.-M., Hsu, M.-H., and Wang, E.T. 2006. "Understanding Knowledge Sharing in Virtual Communities: An Integration of Social Capital and Social Cognitive Theories," *Decision Support Systems* (42:3), pp. 1872-1888.

Chughtai, H., Myers, M.D., Young, A.G., Borsa, T., Cardo, V., Demirkol, Ö., Morgan, C., Morton, S., Prior, C., and Wilkin, J. 2020. "Demarginalizing Interdisciplinarity in Is Research: Interdisciplinary Research in Marginalization," *Communications of the Association for Information Systems* (46:1), p. 13.

Holsapple, C.W., Iyengar, D., Jin, H., and Rao, S. 2008. "Parameters for Software Piracy Research," *Information Society* (24:4), pp. 199-218.

Jiang, L.A., Zhu, N., Yang, Z., Xu, S., and Jun, M. 2018. "The Relationships between Distance Factors and International Collaborative Research Outcomes: A Bibliometric Examination," *Journal of Informetrics* (12:3), pp. 618-630.

Liang, Z., and Yan, Z. 2005. "Software Piracy among College Students: A Comprehensive Review of Contributing Factors, Underlying Processes, and Tackling Strategies," *Journal of Educational Computing Research* (33:2), pp. 115-140.

McDermott, R., and Hatemi, P.K. 2010. "Emerging Models of Collaboration in Political Science: Changes, Benefits, and Challenges," *PS: Political Science & Politics* (43:1), pp. 49-58.

Nunamaker, J.F., Twyman, N.W., Giboney, J.S., and Briggs, R.O. 2017. "Creating High-Value Real-World Impact through Systematic Programs of Research," *MIS Quarterly* (41:2), pp. 335-351.

Nunkoo, R., Sharma, A., Rana, N.P., Dwivedi, Y.K., and Sunnassee, V.A. 2021. "Advancing Sustainable Development Goals through Interdisciplinarity in Sustainable Tourism Research," *Journal of Sustainable Tourism*), pp. 1-25.

O'Sullivan, P.S., Stoddard, H.A., and Kalishman, S. 2010. "Collaborative Research in Medical Education: A Discussion of Theory and Practice," *Medical education* (44:12), pp. 1175-1184.

Oh, W., Choi, J.N., and Kim, K. 2005. "Coauthorship Dynamics and Knowledge Capital: The Patterns of Cross-Disciplinary Collaboration in Information Systems Research," *Journal of Management Information Systems* (22:3), pp. 266-292.

Ou, S., Khoo, C.S.-G., and Goh, D.H. 2008. "Design and Development of a Concept-Based Multi-Document Summarization System for Research Abstracts," *Journal of Information Science* (34:3), pp. 308-326.

Parmiggiani, E., Teracino, E., Huysman, M., Jones, M., Mueller, B., and Mikalsen, M. 2020. "Oasis 2019 Panel Report: A Glimpse at the "Post-Digital"," *Communications of the Association for Information Systems* (47:1), p. 29.

Roschelle, J., and Teasley, S.D. 1995. "The Construction of Shared Knowledge in Collaborative Problem Solving," in *In Computer Supported Collaborative Learning*. Berlin, Heidelberg: Springer, pp. 69-97.

Siponen, M.T., and Vartiainen, T. 2004. "Unauthorized Copying of Software and Levels of Moral Development: A Literature Analysis and Its Implications for Research and Practice," *Information Systems Journal* (14:4), pp. 387-407.

Tarafdar, M., and Davison, R.M. 2018. "Research in Information Systems: Intra-Disciplinary and Inter-Disciplinary Approaches," *Journal of the Association for Information Systems* (19:6), p. 2.

Teston, G. 2008. "Software Piracy among Technology Education Students: Investigating Property Rights in a Culture of Innovation," *Journal of Technology Education* (20:1), Fall2008, pp. 65-77.

Tsai, W., and Ghoshal, S. 1998. "Social Capital and Value Creation: The Role of Intrafirm Networks," *Academy of Management Journal* (41:4), pp. 464-476.

Wasko, M.M., and Faraj, S. 2005. "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *MIS Quarterly* (29:1), p. 35.

Willison, R., and Siponen, M. 2008. "Software Piracy: Original Insights from a Criminological Perspective," *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, pp. 266-266.