2016

Increasing Article Findability Online: The Four C's of Search Engine Optimization

Taryn Marks, *Fredric G. Levin College of Law*
Avery Le, *Fredric G. Levin College of Law*

INCREASING ARTICLE FINDABILITY ONLINE: THE FOUR C'S OF SEARCH ENGINE OPTIMIZATION

Taryn Marks & Avery Le

Reference Librarians

Lawton Chiles Legal Information Center

Fredric G. Levin College of Law

University of Florida

PO Box 117628

Gainesville, FL 32611-7628

352-273-0719 (Marks); 352-273-0712 (Le)

tlmarks@law.ufl.edu; averyle@law.ufl.edu

INCREASING ARTICLE FINDABILITY ONLINE: THE FOUR C'S OF SEARCH ENGINE OPTIMIZATION

Abstract

As researchers increasingly conduct legal research only online, an author must learn the essential skill of ensuring that her article is findable and among the top-ranked results in a search. This article will highlight four search engine optimization best practices to apply to legal scholarship: creating effective titles, abstracts, and metadata; marketing to multiple disciplines; cross-posting to multiple locations; and converting to searchable PDFs, and how to implement each practice to best impact Google Scholar and other search engines' algorithms.

INCREASING ARTICLE FINDABILITY ONLINE: THE FOUR C'S OF SEARCH ENGINE OPTIMIZATION

Table of Contents

I.   Introduction

Promoting scholarship online is difficult. When an author[1] posts her article on the internet, she pits her article against the millions of other articles already available online. In order for an author's article to rise above the masses of other articles, she must actively work to promote her scholarship online. Such promotion is particularly important for an author in a tenured or tenure-track position, where citation counts, impact factors, and online recognition have become increasingly important in the tenure and promotion process, both before and after an author achieves tenure.[2]

To best increase an article's visibility online, an author must practice search engine optimization. To do this, an author should learn how search engine optimization works, how online searches can influence citation counts and impact factors, and how certain techniques can promote the findability of her scholarship

[1] Throughout this article, we use the term "author" to refer to someone who produces academic research publications, most likely for the purposes of tenure.
[2] Emilio Delgardo López-Cózar et al., *The Google Scholar Experiment: How to Index False Papers and Manipulate Bibliometric Indicators*, 65 J. ASS'N INFO. SCI. & TECH. 446, 446 (2014) (noting that "researchers have to respond to evermore demanding pressures to demonstrate their impact in order to obtain research funding or to progress in their academic career, especially in fields of the social sciences and humanities"). "Citation count" refers to the number of citations that an author has to her articles; "impact factor" measures an author's total number of articles, citations, and sometimes the quality of the journals in which the article is published. *See, e.g.*, Raj Kumar Pan & Santo Fortunato, *Author Impact Factor: Tracking the Dynamics of Individual Scientific Input*, SCI. REP. (May 12, 2014), http://www.nature.com/articles/srep04880.

online. To help educate authors who are unfamiliar with search engine optimization, we have developed four "best practices" that we believe best promote scholarship online. Before delving into a discussion of those four best practices, we first clarify how search engines work, and how search engine optimization takes advantage of those search engine processes.

We first explain how search engines work. Because researchers increasingly use Google Scholar to find relevant research, we focus our analysis on the Google Scholar search algorithm and on how researchers use Google Scholar to find information. Many other databases and search engines model their algorithms on the Google algorithm, making an understanding of its underlying functionality even more useful to an author.[3] Once an author understands the likely factors considered by Google Scholar when a researcher conducts a search,[4] the author can use that knowledge to increase an article's findability.

We then explain how we extrapolated the four best practices for search engine optimization. Although we focused our efforts on how to identify practices that would maximize citation counts and impact factors—the metrics that are most important to law professors—we believe the underlying understanding of search

---

[3] In the law field, both WestlawNext and Lexis Advance mimic Google's single search bar. *See infra*, Part II.A.

[4] Unfortunately, Google does not publicize its search algorithm, so we rely on studies that reverse-engineered the Google Scholar algorithm and our own observations of the search engine. *See infra*, Part II.C.ii.

engines and search engine optimization can be applied with equal success across multiple different academic fields and professional disciplines.

Finally, we explain the four best practices and justify why these four are the best practices. We refer to the best practices as "The Four C's of Search Engine Optimization." They are: (1) create effective titles, abstracts, and metadata; (2) cross-discipline post scholarship; (3) cross-post to multiple web locations; and (4) convert works to user-friendly, searchable PDFs. We then conclude, and in doing so also propose additional research projects that would further expand on our own best practices in search engine optimization.

II.   The New Research Paradigm: Google Scholar's Algorithm and its Impact on Research

Google Scholar has taken the academic world by storm, such that researchers almost always start their research in Google Scholar and the habits that researchers develop when using Google Scholar translates across any other database they use.

A. *Researchers Start with and Prefer Google Scholar*

An author needs to understand Google Scholar and how Google Scholar works because both students and academic researchers often start (and too

frequently complete) their research in Google Scholar[5] or even in Google itself.[6]

Students and researchers, already familiar with the Google platform, like Google

Scholar's simple, straightforward search box; more importantly, they like how

easy Google Scholar is to use and how quickly they can find the information they

seek.[7] Because of this "Google fluency," students and academic researchers often

turn to Google Scholar to conduct their initial scholarly searches,[8] instead of using

library catalogs and other academic databases such as Ebsco or ProQuest.[9] Users

in one study located an article using Google Scholar almost 14 times more

frequently than using the library catalog.[10] Even among legal researchers, Google

Scholar has become a preferred search engine.[11]

As a result of Google Scholar's success and popularity, students and

researchers expect (or at least desire) databases other than Google to be Google-

like in how those databases search and present information. Databases have

---

[5] Gail Herrera, *Google Scholar Users and User Behaviors: An Exploratory Study*, 72 C. & RES. LIBR. 316, 318, 319 (2011) (noting that "[Google Scholar] [i]s a good starting place for undergraduate research projects" and that there is "a general adoption of Google among students and researchers alike").

[6] Jillian R. Griffiths & Peter Brophy, *Student Searching Behavior and the Web: Use of Academic Resources and Google*, 53 LIBR. TRENDS 539, 548 (2005).

[7] Herrera, *supra* note 5, at 317–18.

[8] *Id.* at 318–19; *see also* Jöran Beel & Bela Gipp, *Google Scholar's Ranking Algorithm: An Introductory Overview*, *in* 1 PROCEEDINGS OF THE 12TH INTERNATIONAL CONFERENCE ON SCIENTOMETRICS AND INFORMETRICS 230, 230 (2009) (discussing the influence on Google Scholar on the academic scientific community) [hereinafter Beel & Gipp I].

[9] Griffiths & Brophy, *supra* note 6, at 546.

[10] Herrera, *supra* note 5, at 327.

[11] Ashley Krenelka Chase, *Making the Most of Free Legal Research: A Selected Annotated Bibliography*, 28 REFERENCE REV., no. 3, 2014, at 7.

responded to such user preferences by revamping their platforms and search

algorithms to meet researchers' demands.[12] Google Scholar thus plays an

important role in current research techniques, and an author who understands how

Google Scholar works and how researchers use Google Scholar can thus adapt

and apply her knowledge across a multitude of databases.

B. Google Scholar Users won't Discover an Article Past the First Page

We know that researchers default to Google Scholar for academic research;

now we explore how researchers use Google Scholar to best understand how to

optimize scholarship for that type of searching.

Based on eye-tracking analysis, studies determined that researchers

infrequently go past the "page break" of search results,[13] and almost never click

beyond the first page of search results.[14] In Google Scholar, a typical page break

occurs at about the sixth article in the list of search results; a typical first page

includes approximately ten articles. For an article to have the greatest likelihood

of being found by a researcher, the article must appear at least within the top ten

---

[12] *See, e.g.*, Jill Schachner Chanen, *Wired*, A.B.A. J., Feb. 2010, at 35, 37–38 ("[T]he industry's new products will look very much like the Google-ization of legal research.").

[13] Jaewon Kim, et al., *Eye-Tracking Analysis of User Behavior in WWW Search*, in PROCEEDINGS OF THE 27TH ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 479 (2004). The page break is the spot on the screen where a person must scroll down to see the rest of the results.

[14] Andrew Asher, et al., *Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources*, 74 C. & RES. LIBR. 464, 474 (2013).

articles returned by a Google Scholar search; otherwise, the article is unlikely to be seen.

Understanding how Google Scholar researchers use Google Scholar applies to more than just Google Scholar, however. "[P]atterns of use of electronic information systems become habitual,"[15] so that once a researcher establishes the habit of skimming only the first six articles returned from an online search or of looking only at the first page of results, she will maintain that habit across every database and search platform she uses. As more databases imitate Google, the researcher's habitual research techniques become ingrained. This habit even translates to legal databases such as Lexis Advance or WestlawNext (two of the most commonly used databases for legal researchers)—each of which endeavor to be more "Google-like" in their respective search functionality. When researchers use either Lexis or Westlaw to locate an article, the researcher likely sees only those articles near the top of the search results. As such, to maximize the chances that a researcher will notice her article, an author must effectively use search engine optimization tools to push her article to the top of research results, whether in Google Scholar or in any other database.

---

[15] Griffiths & Brophy, *supra* note 6, at 543.

C. Google Scholar and the Google Scholar Algorithm

Before we explore the Google Scholar algorithm, we explain what we know about Google Scholar and its content. An author seeking to grasp how Google Scholar ranks search results and how to increase the ranking of her article must first understand what Google Scholar is.

i. <u>What Google Scholar Searches</u>

Google Scholar searches just a small subset of the internet, and does not draw from websites typified by a Google search, such as Wikipedia, free encyclopedias, or company webpages. Instead, the Google Scholar algorithm pulls results only from sources that Google Scholar deems "scholarly," a term that Google Scholar does not define,[16] leaving us to extrapolate its meaning. Based on our own observations of Google Scholar search results and Google Scholar's examples of the websites it searches,[17] we know that Google Scholar searches and indexes at least some of the articles from the following databases:[18]

- HeinOnline
- SciELO
- ProQuest
- Ebsco

- JSTOR
- SSRN
- Wiley
- Elsevier

---

[16] *See, e.g.*, About, GOOGLE SCHOLAR, https://scholar.google.com/intl/en/scholar/about.html (last visited Feb. 24, 2016).

[17] Inclusion Guidelines for Webmasters, GOOGLE SCHOLAR, https://scholar.google.com/intl/en/scholar/inclusion.html (last visited Feb. 18, 2016).

[18] This list is not exhaustive, as we know that Google Scholar searches many additional databases.

- bePress (and other institutional repository hosting sites such as DSpace)

- Sage

- LexisNexis

Additionally, if an author posts a PDF on the publications page of a .edu website, with both a title and author at the top of the first page and either a list of references or a bibliography somewhere in the PDF, that author's paper will also be indexed by Google Scholar.[19]

While we know the general outline of the sources that Google Scholar searches, we do not know its scope, such as the percentage of articles within the databases that Google Scholar accesses and indexes, and the parameters of that percentage; or Google Scholar's depth, such as whether the algorithm searches all of the databases every time it runs a search.

We also know that Google Scholar indexing an article and presenting that article in search results does not equate to researchers having access to the full text of that article. Google Scholar has arrangements with most of the databases listed above, arrangements that allow the Google Scholar bots to index and list the databases' articles in its search results. But almost all of the databases listed above

---

[19] Inclusion Guidelines, *supra* note 17.

are subscription databases, so that only those researchers who have a subscription to those databases can read the full text of those articles.

Libraries have addressed some accessibility concerns. Google Scholar does contract with local libraries to crawl library catalogs and will link a researcher to the library catalog if the article might be found in that library. Universities with IP-authenticated access to subscription databases also allow Google Scholar to seamlessly link researchers from the results of a Google Scholar search to the full text of articles behind paywalls. But for many researchers, Google Scholar search results provide only a tantalizing glimpse of potentially useful articles. An author should request that an article be posted in one of the databases Google Scholar indexes; an author should also post an article on a faculty webpage, SSRN, or institutional repository, to provide access for those who otherwise would not be able to use or cite the article.

An author seeking to increase the visibility of scholarship online must consider the sources Google Scholar draws from and the availability of an article to a researcher using Google Scholar. In these ways, an author can take advantage of her knowledge of what Google Scholar searches; she can further take advantage of Google Scholar once she also knows how Google Scholar works.

ii. How the Google Scholar Algorithm Works

In addition to understanding what Google Scholar searches, an author must understand how the algorithm searches for and then ranks the articles listed in

search results. To achieve the best optimization of an article, an author must work with our understanding of the Google Scholar algorithm's methodology and ranking system.

Unfortunately, we do not know the details of the Google Scholar algorithm, as Google refuses to divulge the secret of its methodology and shares only a few vague details with the public.[20] But, since Google unveiled Google Scholar in November 2004,[21] computer scientists and others have reverse engineered the algorithm and extrapolated several factors that they believe the Google Scholar algorithm measures and utilizes to generate its search result rankings.[22] Jöran Beel and Bela Gipp examined the algorithm using several different methodologies, focusing on three variables: a basic keyword search, citation count, and the age of an article.[23] From these studies, Beel and Gipp developed the following theories:

---

[20] Currently, Google Scholar says only that "Google Scholar aims to rank documents the way researchers do, weighing the full text of each document, where it was published, who it was written by, as well as how often and how recently it has been cited in other scholarly literature." About, *supra* note 16.

[21] Judit Bar-Ilan, *Which h-index? – A comparison of WoS, Scopus and Google Scholar*, 257, 258 (2008).

[22] *See, e.g.*, Beel & Gipp I, *supra* note 8; Jöran Beel & Bela Gipp, *Google Scholar's Ranking Algorithm: The Impact of Citation Counts (An Empirical Study)*, in THIRD INTERNATIONAL CONFERENCE ON RESEARCH CHALLENGES IN INFORMATION SCIENCE, 439 (2009) [hereinafter Beel & Gipp II]. *See, e.g.*, Jöran Beel, et al., *Academic Search Engine Optimization*, 41 J. SCHOLARLY PUB. 176 (2010); Péter Jacsó, *Google Scholar: The Pros and the Cons*, 29 ONLINE INFO. REV. 208 (2005).

[23] Beel & Gipp I, *supra* note 8, at 230; Beel & Gipp II, *supra* note 22, at 439; Jöran Beel & Bela Gipp, *Google Scholar's Ranking Algorithm: The Impact of Articles' Age (An Empirical Study)*, in SIXTH ANNUAL CONFERENCE ON INFORMATION TECHNOLOGY: NEW GENERATIONS, 160, 160 (2009) [hereinafter Beel & Gipp III].

Factors that <u>directly influenced</u> search result ranking of an article	Factors that <u>did not directly influence</u> search result ranking of an article
Citation count of the article[24]	Number of times a search term appears in the article[25]
Exact search terms appear in the article[26]	Synonyms of search terms appear in the article[27]
Search terms appear in the article's title[28]	Publication date of the article[29]

Beel and Gipp also determined that during a title field search, the algorithm *always* factors the citation count of the article into the ranking of results. During a full-text search, however, the algorithm *usually* factors the citation count, but not always.[30] We speculate that when a researcher searches for keywords in the title field, search results tend to be more limited, so the algorithm assumes that the researcher would prefer an article with a higher reputation. When a researcher does a full-text keyword search, however, the number of possible articles

---

[24] Beel & Gipp I, *supra* note 8, at 233; Beel & Gipp, *supra* note 22, at 242–43.
[25] Beel & Gipp I, *supra* note 8, at 234, 235.
[26] *Id.* at 234.
[27] *Id.*
[28] *Id.* at 235.
[29] Beel & Gipp III, *supra* note 23, at 164.
[30] Beel & Gipp II, *supra* note 22, at 444.

expands, so the algorithm will usually rank articles by citation count, but not always.[31]

Additionally, the publication date of an article does not directly influence the search result ranking of that article, but articles in the top-ranked positions "are on average older than articles" in the lower-ranked positions.[32] We hypothesize that this occurs because an article's publication date directly influences that article's citation count: an older article, published last year, will likely have a higher citation count than an article published yesterday, simply because of time.

Finally, Beel and Gipp demonstrated that the Google Scholar algorithm changes depending on whether the researcher uses the title, full text, cited by, or related article search function.[33] We think that perhaps Google Scholar "reads" a search string and determines an article's relevance differently depending on how the researcher constructs a search. A title search for "originalism" likely indicates to the algorithm that the researcher wants articles in which originalism is the exclusive topic of the article, while a full text search for "originalism" may indicate to the algorithm that the researcher would like articles that discuss originalism, not articles in which originalism is the exclusive topic of the article. This may also be one of the reasons why Google Scholar does not seem to

---

[31] For example, if an article has keywords both in its title and in its text, the algorithm may rank that article higher than an article with a higher citation count but keywords only in its text.
[32] Beel & Gipp III, *supra* note 23, at 163.
[33] Beel & Gipp I, *supra* note 8, at 235.

consider the number of times that a search term appears in the full text of an article.

All of this reverse engineering and understanding of the Google Scholar algorithm, though, only helps an author who implements these techniques, such as the four best practices.

III.   The Four Best Practices

We originally began identifying best practices to improve online scholarship visibility for a workshop that we presented to the University of Florida law faculty. To delineate the best practices, we researched how search engines such as Google Scholar function, but we also pooled our experiential knowledge and observations of online searches. Additionally, we both have expertise in the creation and management of institutional repositories. At the University of Chicago's law library, Taryn worked on defining and choosing the metadata most useful to a historical institutional repository. In her work at Levin College of Law, Avery frequently came across articles that contained weak and inaccurate metadata. Avery's experience managing repositories provided examples of good and bad metadata, and of how bad metadata could prevent information from being

found.[34] Our technical research[35] and our experience led us to the "Four Cs" of search engine optimization for legal scholarly works.

A. Create Effective Titles, Abstracts, and Metadata

The first best practice is to carefully craft the title of any article, write a short abstract filled with keywords, and verify that an article's metadata is correct.

i. Write an Effective Title

First, an author should write an effective title. A title serves as the first point of contact with a researcher and acts as one of the key components of an article's embedded metadata. Researchers who skim just the first page of search results quickly judge an article based solely on its title. As such, an author must carefully choose words that effectively summarize the contents of the article in only a brief snippet, and should incorporate important keywords into the title to make the article more findable to researchers conducting a search.[36]

Ideally, the title will appeal to readers, so an author should create a smart, witty title that does not detract from the article's content: a challenging endeavor. Practically speaking, researchers are more likely to click on an article with a clear and accurate title that concisely describes the article's subject matter or main

---

[34] For example, in one of the articles, the named author was the research assistant of one of the professor's previous co-authors.
[35] *See supra*, Part II.B., C.
[36] *See supra*, Part II.C.ii.

thesis, than they are to click on an article with an abstract, obtuse title. So

although a creative title may attract a researcher's attention, the title's catchiness,

without appropriate keywords, will likely push that article lower in a search result

ranking than will a non-pithy title that has meaningful keywords. A cleverly titled

article loses its value if researchers will not find the article at the top of their

search results, so an author should favor information and keywords over

wittiness.[37]

An author must also consider the length of the title. A title should balance

being catchy and informative, yet avoid the risk of being misleading, verbose,

curt, or exhaustive to the point of overwhelming the researcher. Including a

subtitle can allow an author to grab a researcher's attention with the title and then

to fully convey the substance of the article within the subtitle. When writing a

subtitle, keyword positioning is critical. In Google Scholar (and some other

databases), search results usually include only the first seven or eight words of a

title. Because of this quirk, an author should put the most important content of the

title first, then insert a colon (a frequently-used theme in law) for the clever,

creative part of the title. By placing the explanatory part of the title first, the

researcher will be able to readily decipher the subject of an article, instead of

---

[37] *See supra*, Part II.C.ii.

having to guess at the meaning of the clever and creative part of the title. Title transparency matters.

The most important practice with respect to article titles is to insert keywords directly into the titles. As Gipp and Beel revealed, keywords in the title influence an article's ranking far more than the keywords in the article itself.[38] For example, if an article's main focus is online data privacy, the author should highlight specific keywords such as "data privacy" and "online data protection" in the article's title. Although the title "The Right to Be Forgotten" may describe the article's contents and correlate with data privacy, the title may be misinterpreted if the researcher, especially a researcher outside of the law field, is unfamiliar with the concept of *the right to be forgotten*.[39] Adding a colon and several more words to the title, such as *The Right to be Forgotten: Protecting Data Privacy in the Internet Era*, helps to explain the right to be forgotten to those researchers without that knowledge. Even better would be *Protecting Data Privacy in the Internet Era: Asserting the Right to be Forgotten*, which would place the critical keywords at the front of the title.

When an article's title matches the keywords in a researcher's search string, that article will likely rise in search result rankings and will be more likely to be

---

[38] *See* Beel & Gipp I, *supra* note 8, at 234, 235.

[39] We borrowed this title and example from Robert Kirk Walker, Note, *The Right to be Forgotten*, 64 HASTINGS L.J. 257 (2012). The right to be forgotten is the right to have one's information deleted from the internet (especially after death) and to have complete privacy from internet search engines.

read, used, and cited. Additionally, by continually connecting broader topics (online data privacy) with an important subset of that topic (the right to be forgotten), search engine algorithms may start to recognize the connection, and offer researchers the suggestion of "right to be forgotten" when the researcher searches "online data privacy."

Creating an effective title, with clear keywords that accurately convey the subject of the article is just the first step of this best practice. Next, an author must consider the abstract.

ii. Craft an Effective Abstract

After an author catches the researcher's attention with the title, she must then draw the researcher in with the abstract. As such, an author should write short, accurate abstracts that contain several keywords (ideally at the beginning of the abstract).

An abstract should provide more detail than a title, expanding upon what the researcher learned from the title, but the abstract cannot be so long that it loses concision and clarity. The first lines of the abstract have significant weight on the researcher's decision to open the article, which means the repetitive use of important keywords within the abstract, particularly at the beginning, ensure that researchers will quickly grasp the article's subject matter and will open those that are most relevant or interesting to the researcher. In some search engines, such as Google Scholar, the first few lines of the abstract appear on the search results

page. Abstracts that appear on the search results page greatly impact an article's findability because the researcher does not have to click into another document to determine if the article is worth reading. Because the researcher can only see the first few lines of the abstract, those first lines must convince the researcher to click into the article.

A word of warning, however: an author should avoid trying to "game" the system, by plastering a certain keyword repetitively in the abstract. This technique will fail to get the researcher's attention and will not result in a higher ranking in the search algorithm because the system will recognize that the abstract is "fake" and that the author is attempting to exaggerate certain elements to trick the system, which can result in the search algorithm completely removing an article from a search. Construct effective, clear abstracts that accurately convey the contents of the article to avoid these potential downfalls.

iii. Ensure Effective Metadata

The final component of the first best practice is the metadata. If the title and abstract are the protagonists on the main stage of internet search results, then consider metadata as the behind-the-scenes production crew helping the show come together. Metadata contains descriptive information embedded into an article that reflects its contents. Researchers will not necessarily see metadata on the screen; on the back end, though, metadata is the essential component for electronic transmission of information. Computers see metadata like a blueprint of

contents, using metadata to calculate and extract information for the search results.

Search engines rely heavily on underlying metadata when indexing articles and determining relevancy, so this information must be as complete and accurate as possible. Although many programs automatically input metadata when a document is first created, the quality and content of that metadata varies based on several factors that are not germane to the content of the article: e.g., the program used to create the metadata or the format of the document. An author should always check the accuracy of the metadata before publishing an article on the internet because the consequences of incorrect metadata can be dire. An article that incorrectly lists the title or keywords (such as "Second Amendment" instead of "First Amendment") will be hidden among the masses of search results because the search algorithm will not recognize the article's relevance to a researcher's search string. The small amount of time an author devotes to checking for correct metadata is slight compared to the potential consequences of inaccurate metadata, so verifying the metadata is a must for authors.

Below, we discuss in detail the metadata underlying a PDF, as we assume that almost all articles are being posted in that format.[40] A PDF has three key metadata pieces: the title, the author, and the keywords.[41]

An author's easiest metadata field is the title field: an author should simply confirm that the title of the article in the metadata is correct. If not, an author corrects the metadata to reflect the right title.

The second field is an author's name. Entering an author's name, though, can be deceptively tricky because of the different variations an author may choose, such as middle initials, middle names, maiden names, etc. Names are important because, over time, an author may come to be known as an expert in a specific field, so a researcher may try to search for articles using a specific name. An author also establishes a scholarly presence online, so that an author who is well known in a certain field is more likely to be cited by those researching in that field. If an author sometimes goes by John R. Smith, sometimes by John Smith, and sometimes by John Roe Smith, how is a researcher (or a search engine) to know whether those three names represent the same author, two separate authors, or three separate authors?

---

[40] We encourage all articles to be posted as PDFs; the format is more stable than Word documents online and can be more difficult to manipulate.

[41] For instructions on how to access the metadata in a PDF and how to check the title, author, and keyword metadata fields, please see the Appendix

Whether an author chooses to use a middle initial or middle name, the author must be consistent, and all author fields in all articles that an author posts online should have the same name. This ensures that researchers find the specific author that they are looking for and can help increase search result rankings because the search engine will be able to attribute all citations to one author, rather than splitting up citations because the search engine sees John Smith and John R. Smith as different people.

The last field of important metadata is the keyword field. As with the title and author name of an article, the keywords in the metadata must also be correct. Enter keywords that accurately reflect the content of the article and its area of law, similar to the contents of the title and abstract. We estimate that ten to twelve keywords is a good number; the metadata field needs enough to accurately convey the contents of the article, but with too many keywords the article will come up in results for which it is not relevant, discouraging researchers and potentially harming an author's online reputation.

The first best practice tells an author to write a clear, accurate title, with a short, sweet abstract with keywords, and to ensure the metadata underlying an article is correct. Doing so can help an author increase the chances that a researcher will find an author's article and will click into that article.

B.   Cross-Discipline Posting

The second best practice is to market an article across multiple disciplines and under multiple sub-disciplines. Many law articles discuss more than law, and cross into other disciplines such as criminal justice, gender studies, or economics. When an author writes an article that deals with both subject matters, the author should post the article in law databases, as well as in economics databases, gender databases, and any cross-discipline databases, such as a law and economics database. An author must be careful, however, to ensure that she is not posting in databases that are completely unrelated to the topic of the article: search engines can pick up on when an author is gaming the system, and it can also reduce an author's reputation in the academic community if it becomes known that an author frequently exaggerates the subject matter of her articles. The credibility of the work may be contingent on an author's expertise in the field, so an author should maintain her reputation and stay accurate with the discipline selection.

The key is to choose pertinent and relatable disciplines. An author should think about which disciplines form a broad umbrella that encompass a specific issue and topic, even if it does not coincide directly with the area of law. For example, consider the Legal Scholarship Network series on SSRN, which hosts a broad range of discrete topics, organized under big umbrella topics.[42] By posting

---

[42] Legal Scholarship Network eJournal Taxonomy, SSRN, http://ssrn.com/en/index.cfm/lsn/lsn-ejournals/ (last visited Feb. 28, 2016).

across multiple disciplines, an author gains more exposure for her work and disperses it to a wider range of researchers who may have different academic backgrounds other than law. This effort will can maximize an author's readership ten-fold.

By posting across multiple disciplines, an author can also attract different audiences and can catch those who search by broad topic. Additionally, if an author inserts the cross-discipline as a keyword in the metadata of the article, it will increase the chances that a researcher will find the article. Much research today is being conducted across disciplines, so an author who can capture multiple markets increases her chances of being cited and of being recognized as an expert in several, related fields.

C. Cross-Post in Multiple Locations

The third best practice is to post an article (or the draft of an article) in multiple different places. Posting to several different locations helps an author reach a wider range of potential researchers. An author should post to a faculty biography page on a school's website; to SSRN or ResearchGate; to her LinkedIn profile; to her host institution's Institutional Repository or bePress Selected Works page; to any blogs, Twitter feeds, Facebook pages, etc. that an author has; and to any personal or professional websites that an author maintains. Of these different sites, posting to a faculty biography page, SSRN, and a host institution's

Institutional Repository, are the most important because of the credibility these sites give to an author's work, as opposed to an unverified work on a privately-hosted webpage. By doing so, an author can increase the article's ranking in Google Scholar, as there is some evidence that Google considers the number of disparate places in which an article is located as a factor in ranking that article.

By posting across multiple locations, an author has the opportunity to catch disparate researchers. Different researchers regularly search SSRN, for example, than LinkedIn, simply because of the familiarity they have with SSRN and their knowledge of its use by academics as a scholarship repository. Other researchers may visit an author's LinkedIn page for networking purposes, but may also be intrigued by the author's publications, especially if an author is a known expert in a particular subject area. By posting to both SSRN and LinkedIn, an author can attract both of those audiences, increasing the chance that her scholarship will be read and that the scholarship will be cited.

Posting one's scholarship to multiple locations also helps an author create a strong online reputation. Being mentioned on different websites will reiterate the impact that the article has on the scholarly community, and in turn will enable the article to become a topic of discussion in the academic field. Multiple search results will also translate into more download counts for online statistics.

As with gaming the abstract, be careful about floating duplicate versions of the same publication, which may result in an imbalance of citation counts. To

minimize this potential predicament, an author should link to an original copy that

is hosted on one centralized server, such as SSRN or an institutional repository,

instead of re-uploading the PDF to a new location each time. But by credibly

posting across different websites, an author will increase the findability of her

articles.

D. *Convert the PDF into a Searchable PDF*

The final best practice is to only post PDFs that have been converted into

searchable PDFs, or to OCR (Optical Character Recognition) the PDF.[43] PDFs are

often posted to the internet without being converted into text. This effectively

renders the PDF as an image, preventing a search engine from "reading" the

individual words in the article. By converting the PDF to a searchable PDF, an

author transforms the image into a readable document. A search engine can now

"read" the article word by word, and bring it up in a search result if it is relevant

to that search, especially useful for those researchers who rely on keyword

searches within the text. Without readable text, the text of the PDF, even if it

includes relevant keywords that researchers are typing into Google Scholar, may

go unrecognized, limiting the article's opportunities to be found by the search

engine.

---

[43] The assumption, of course, is that all authors are already posting articles in PDF format, not in
any other format (Word, Works, RTF, etc.). Authors not posting articles in PDF format should
immediately change that practice.

An OCR'd article complements the accurate metadata encrypted into the article's back end. Posting only searchable PDFs ensures that even if the metadata is missing vital information, the article will still be discoverable in search results. Additionally, a researcher can search through an OCR'd PDF to find specific sections of scholarship, increasing the chances that the researcher may cite to that specific section.

To even further increase the likelihood that an article is on the first page of results, an author should post only machine-readable PDFs.

IV.   Selling the Best Practices

Although we believe that the best practices can impact an author's reputation and citation count, and we purposely designed them to be simple and quick to implement, it can be frustrating to even inform others about the best practices, much less convince them that they should implement them. We know many authors who are aware of the best practices but who continue to publish without checking metadata or converting to a searchable PDF. Our institutional repository's staff must double-check all information before uploading an article. A simple scroll through an SSRN eJournal will demonstrate how confusing article titles can be and how many authors still compose witty titles without context.

Part of the best practices, then, must also be marketing the best practices and convincing other authors that the few minutes they take to implement reap enormous rewards. We point out the increased findability of articles, which leads to increased citation counts. In law schools especially, an increased citation count has come to be seen as an objective marker of influence and success, so we can connect the best practices to a specific, practical goal.

Marketing the best practices must also involve disabusing others of possibly inaccurate views of search engine optimization. For example, changing the metadata and OCR'ing an article already posted on SSRN does not restart the download counts for that article. There may be other misinformation about how search engine optimization works floating around. We must keep our ears open to any rumors about the best practices so that we can quickly and efficiently correct any confusion.

As yet, we rely heavily on regularly reminding others of the best practices and offering to walk them through any of them. Working with a dean of faculty development may also be an option, and convincing one of the more active, influential faculty members may be another. Sending faculty this article (or others like it) can be another starting point. We also think it important to implement the best practices for any scholarship you may produce, to lead by example.

V. Conclusion and the Future of Search Engine Optimization

By implementing the four best practices, we believe that an author has the opportunity to increase the chances that her scholarship will be found online and that the scholarship will then be cited in future articles. The best practices, based on a strong foundation of research and real-world experience, are easy to implement, practical, and will likely be successful.

But simply implementing the best practices must only be the first step. Google Scholar and other search engines are bound to change. The way in which we search and find material is equally destined to change. One of the keys to the success of the best practices is that it responds to the current research and the current methods of searching. Today the best practices may serve as useful tools, tomorrow's search engines may change that.

Continuing to research and track how search engine optimization works and the best methods for optimizing research should be a priority. We would be interested to know the impact of search engine optimization on a scholarly article. We would want to do case studies of SSRN, institutional repositories, and Google Scholar to further assess and articulate how those search engines operate and how researchers use those search engines. We would be interested in seeing whether Facebook's "Boost" option for a liked page could be applied to an article that's been cited on Google Scholar.

The four best practices for search engine optimization offer legal scholars the opportunity to increase their visibility to the academic research world. And a greater familiarity with how search engines work and how researchers find articles, along with a curiosity about the future of search engine optimization, means that we will only continue to expand the opportunities to increase that visibility.
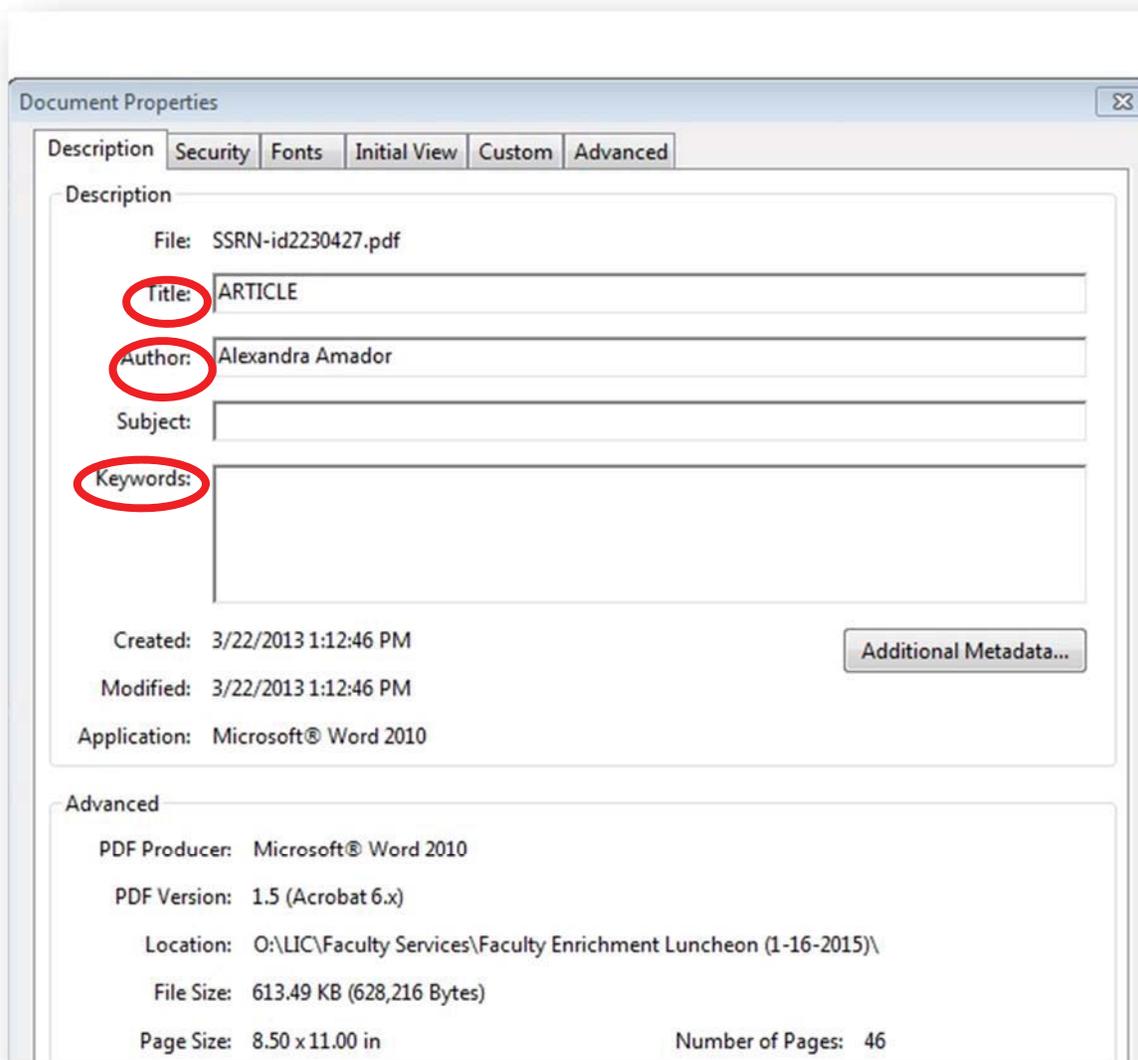
Appendix

Instructions: Best Practice 1: Checking PDF Metadata

1. Open your PDF file in Adobe (either Reader or Acrobat).

2. Click "File" in the top left corner.
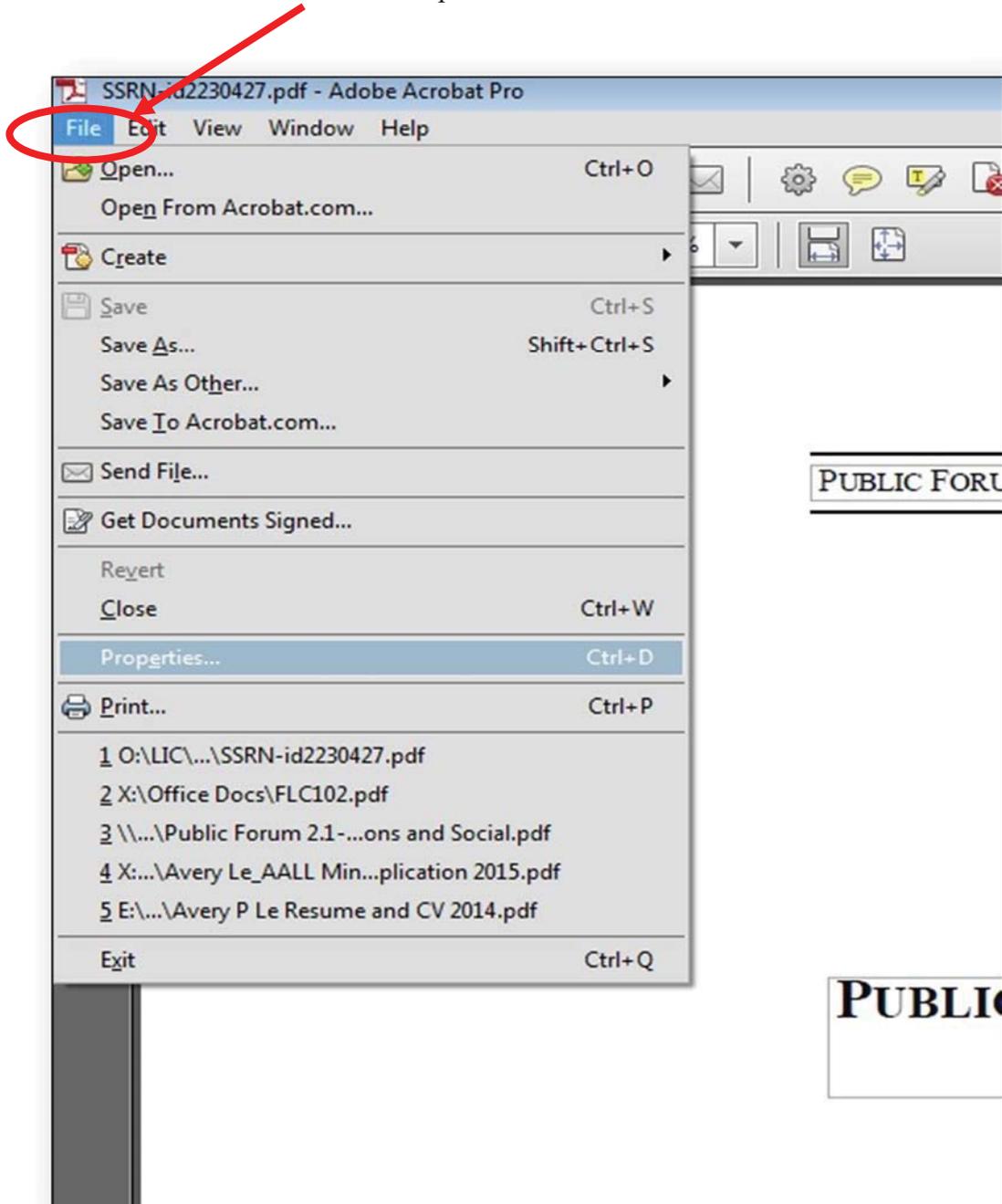
3. Select "Properties" and click on it. You will see the description of the PDF. In the "Title", "Author", and "Keywords" fields, type in the appropriate information – this will be the metadata for the PDF.

4. Once you are satisfied with your data, click the "OK" button to save your changes. The PDF metadata will now be updated and intact for your next upload.

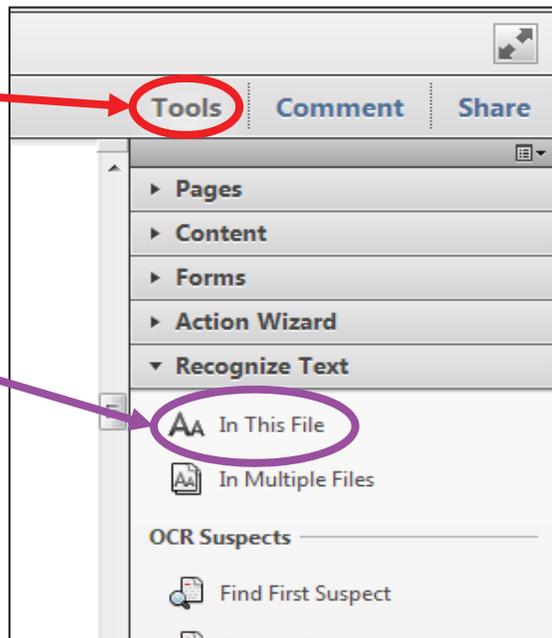**Instructions for Best Practices 4: Convert to Searchable PDF**

1. Open a document in Adobe Acrobat
2. Click ① View
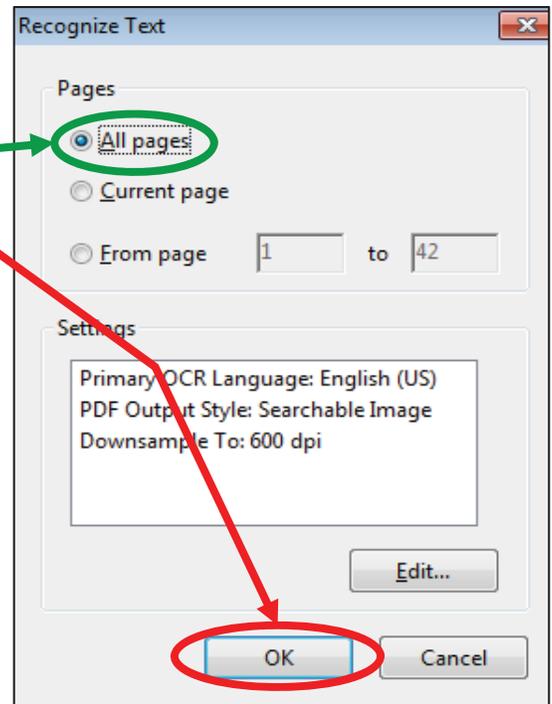   ② Tools
   ③ Recognize Text
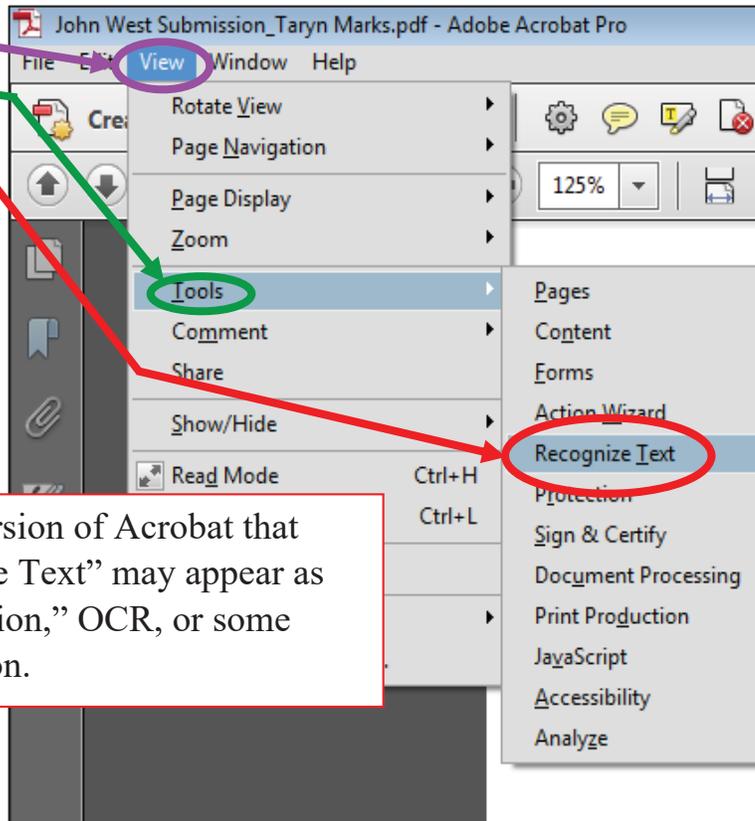
Depending on the version of Acrobat that you have, "Recognize Text" may appear as "OCR Text Recognition," OCR, or some other, similar, iteration.

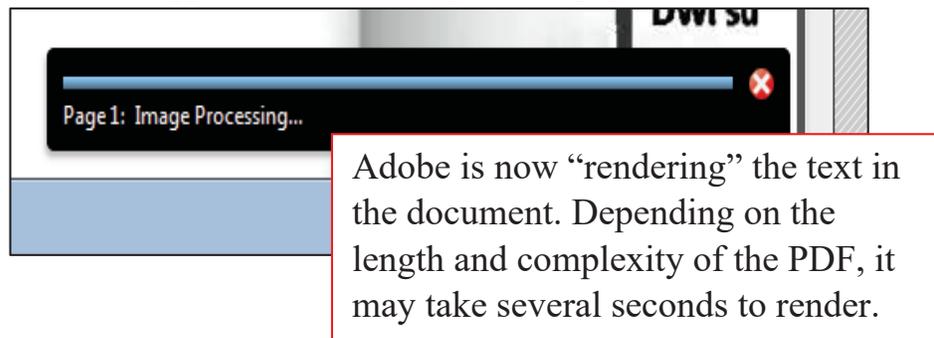3. The Tools box will open to the right of the document:
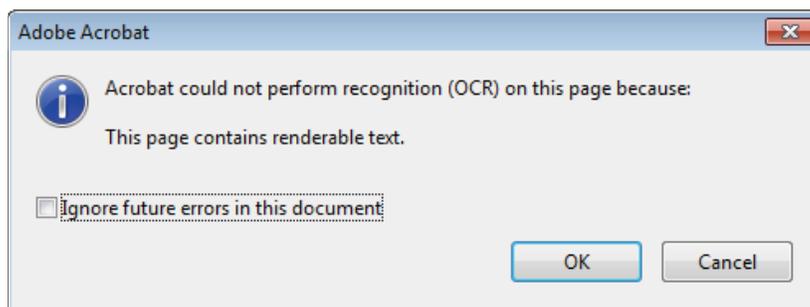
4. Under "Recognize Text," click "In This File."

5. The "Recognize Text" box will pop up.

Confirm "All pages" is marked, then click "OK."

Recognize Text

Pages
- ⦿ All pages
- ○ Current page
- ○ From page [1] to [42]

Settings

Primary OCR Language: English (US)
PDF Output Style: Searchable Image
Downsample To: 600 dpi

Edit...

OK   Cancel

A black dialogue box will appear at the bottom right of the page as the PDF is being OCR'd:

Page 1: Image Processing...

Adobe is now "rendering" the text in the document. Depending on the length and complexity of the PDF, it may take several seconds to render.

If, rather than the black box, the renderable text box pops up after clicking OK:

Adobe Acrobat

ⓘ Acrobat could not perform recognition (OCR) on this page because:

This page contains renderable text.

☐ Ignore future errors in this document

OK   Cancel

then the document has already been converted to searchable text and no further steps need to be taken.