A Comparative Analysis of Research Literature Resources

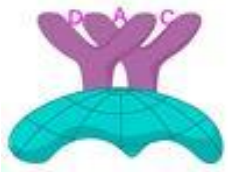
An Interactive Qualifying Project Report
Submitted to the Faculty
of the
WORCESTER POLYTECHNIC INSTITUTE

WPI

in partial fulfillment of the requirements for the
Degree of Bachelor of Science
by:

Lidya Gebremeskel
Nathaniel O'Connor
Oliver Spring
Olivia Steen

Sponsored by:
Dr. Xiaomai Zhou (周晓迈), Vice President
Department of Therapeutic Antibodies, Hangzhou DAC Biotech



Date: 15 December 2016

Professor Gary Pollice, Co-Advisor
Worcester Polytechnic Institute

Professor Stephan Sturm, Co-Advisor
Worcester Polytechnic Institute

Abstract

This report addresses some of the challenges researchers at Hangzhou DAC Biotech face in acquiring full-length articles and staying up to date when conducting literature research. The goal of this project, sponsored by Hangzhou DAC Biotech, is to analyze their current resources and recommend alternatives and changes to improve their literature research. The results of this project show that a combination of literature resources present the best options for researchers to acquire full-length texts. We also implement an NCBI notification system to help Hangzhou DAC Biotech researchers stay up to date with the newest publications.

Acknowledgements

This project would not have been possible without the guidance and support of several individuals that we would like to acknowledge. First and foremost, we would like to thank our sponsor, Dr. Xiaomai Zhou, for arranging and supervising this project, as well as the researchers at Hangzhou DAC Biotech. Second, we would like to thank our project advisors, Professor Gary Pollice and Professor Stephan Sturm, for giving us guidance throughout our proposal and project. We would like to acknowledge Professor Yunhong Shen of Hangzhou Dianzi University and Professor Jennifer Rudolph for organizing the 2016 Hangzhou project center. We thank our HDU buddies for assisting us with translations and introducing us to Chinese culture. Lastly, we would like to thank Paige Neumann, as well as the rest of the WPI library staff, for taking the time to participate in our interviews.

Executive Summary

Cancer is the leading cause of death in China with about two million deaths per year. The prevalence of this disease makes research on cancer treatment methods an important aspect in the field of biotechnology. One promising treatment method on the forefront of the biotechnology industry is the use of Antibody Drug Conjugates (ADCs). ADCs are one of the newest methods of combating tumors in cancer patients by targeting infected cells while leaving healthy tissue untouched. Staying current in this emerging field requires access to the most recent research literature.

Our sponsoring company, Hangzhou DAC Biotech, is researching ADCs and has trouble accessing and staying up to date with current literature in their field. A major hindrance for Hangzhou DAC Biotech is that most research papers lay behind steep paywalls set by publishers. Due to their small budget for subscriptions to research literature resources, Hangzhou DAC Biotech faces difficulty with effectively finding detailed information about new developments in the ADC field.

The goal of this project is to recommend literature resources that allow researchers to access full-length articles in a cost-effective manner. Additionally, we institute a notification system that alerts researchers when new articles pertaining to their research are published.

Current resources in use at Hangzhou DAC Biotech

Through interviews and surveys with Hangzhou DAC Biotech researchers, we determine that the National Center for Biotechnology Information (NCBI) is the most used literature resource because it is specific to their field of study and free to browse. Despite being blocked in China, Google is another popular resource. The need for VPN when using Google makes it unstable and difficult to use. Once an abstract of interest is found on one of these literature resources and the researcher needs the full-length article, most stated that they use personal connections to acquire them. They contact former colleagues or friends associated with institutions with subscriptions that allow access to full-length literature.

Our sponsor stated that Hangzhou DAC Biotech subscribes to ScienceDirect, paying 400 USD per year for a subscription. However, even with this subscription, finding full-length articles is a problem for the researchers.

Potential solutions for Hangzhou DAC Biotech

To help researchers get the detailed information they need, we look into new, cost-effective avenues for accessing full-length papers. In addition to finding avenues for them to access full-length articles, we investigate ways to help researchers stay up to date on newly published articles. We institute a notification system for their favorite literature resource, NCBI.

Upon investigation, we conclude that the best way for Hangzhou DAC Biotech to increase their access to full-length articles is to use a combination of literature resources. We recommend substituting the company's current subscription to the literature database ScienceDirect with a subscription to DeepDyve. We suggest the company use ScienceDirect's free searching capabilities in addition to the free searching capabilities of NCBI.

To help researchers stay up to date on recently published articles, we research notification alert systems. We explore an alert system on their favorite online research literature platform, NCBI. We created accounts for all researchers who requested an alert system and drafted a tutorial detailing how to manage and change the alerts. We recommend continuing the use of this notification system and periodically updating the search alerts to ensure the notifications stay relevant to their current research.

Authorship

Section	Primary Author(s)	Primary Editor(s)
Abstract	Lidya Gebremeskel	Nathaniel O'Connor Olivia Steen
Executive Summary	Olivia Steen	Lidya Gebremeskel Oliver Spring
Chapter 1: Introduction	Oliver Spring	Nathaniel O'Connor Lidya Gebremeskel
Chapter 2: Background		
2.1 Literature Databases	All	All
2.2 Important Factors for Literature Databases	Lidya Gebremeskel Olivia Steen	Nathaniel O'Connor Oliver Spring
2.3 Online Literature Database and Literature Resources	All	All
2.4 Literature Database Sharing Options	Nathaniel O'Connor Oliver Spring	Lidya Gebremeskel Olivia Steen
Chapter 3: Methodology		
3.1 Evaluate Current Methods and Resources	Nathaniel O'Connor Olivia Steen	Lidya Gebremeskel Oliver Spring
3.2 Investigate Alternative Research Resources	Lidya Gebremeskel Oliver Spring	Nathaniel O'Connor Olivia Steen
3.3 Analyze literature databases	Lidya Gebremeskel Olivia Streen	Oliver Spring Nathaniel O'Connor
Chapter 4: Results and Analysis		
4.1 Analysis of Current Methods and Resources	Nathaniel O'Connor Lidya Gebremeskel	Olivia Steen Oliver Spring
4.2 Evaluation of Potential Solutions	Olivia Steen Oliver Spring	Nathaniel O'Connor Lidya Gebremeskel

4.3 Literature Resource Comparison	Nathaniel O'Connor	Olivia Steen Oliver Spring
Chapter 5: Conclusions and Recommendations		
5.1 Conclusion	Olivia Steen Lidya Gebremeskel	Oliver Spring Nathaniel O'Connor
5.2 Limitations	Lidya Gebremeskel Oliver Spring	Nathaniel O'Connor Olivia Steen
Appendices	All	All

Table of Contents

List of Figures

List of Tables

1. Introduction

Cancer is the leading cause of death in China, causing about two million deaths per year (Chen, 2016). The prevalence of this disease makes research on cancer treatment methods an important aspect in the field of biotechnology. Antibody Drug Conjugates (ADCs) are one treatment method on the forefront of the biotech industry. ADCs are one of the newest methods of combating tumors in cancer patients; they target infected cells while leaving healthy tissue untouched (Bakhtiar, 2016). Staying up to date in this emerging field requires access to the most current research literature.

Our sponsor, Hangzhou DAC Biotech, is a small cancer research company with the goal of developing an ADC treatment. Currently, only three ADCs have received the Federal Drug Administration's (FDA) approval (Bakhtiar, 2016). Hangzhou DAC Biotech hopes to push their developing ADCs to clinical trials within three years, with the long-term goal of releasing an affordable ADC drug onto the market (Zhou, X.M., personal communication, November 11, 2016). They require access to full-length publications to achieve this goal. However, due to their small budget for literature resource subscriptions, Hangzhou DAC Biotech faces difficulty with effectively accessing articles about new developments in their field. A full sponsor description is found in Appendix A.

The cost of literature resource subscriptions is a major limitation for smaller companies. The cost of subscriptions can vary depending on the type of user and their intended use. A company can subscribe to a platform which contains many searchable full-length articles, but this can be expensive. Resource sharing between institutions to receive group discounts on subscriptions is another option. If the researchers do not need to access many articles, finding a resource that allows the researchers to pay per article at a reduced cost is another option. We consider all of these factors to choose a cost-effective solution.

The company currently only subscribes to parts of the literature database ScienceDirect in addition to accessing free resources, such as the National Center for Biotechnology Information (NCBI) (Zhou, X.M., personal communication, November 11, 2016). This combination of literature database resources is not enough to obtain all the full-length articles the researchers at Hangzhou DAC Biotech need. ScienceDirect has many full-length articles but is not all-encompassing. NCBI often only grants access to the abstracts of the articles, which must be

purchased individually to view full-text (Zhou, X.M., personal communication, November 11, 2016). This steep paywall hinders the speed and quality of Hangzhou DAC Biotech's research. Due to the increasing amount of different medical research journals and literature databases, we explore other cost-effective options with greater article coverage.

Researchers must be up to date on the latest publications to stay at the top of their field. Because researchers at Hangzhou DAC Biotech often focus on laboratory work, they find it challenging to stay up to date on recently published articles (Zhou, X.M., personal communication, November 11, 2016). Many literature resources ease this strain by offering an alert system that notifies researchers when relevant articles are published. We explore and implement some of these notification systems to help Hangzhou DAC Biotech researchers stay current with relevant publications.

The goal of this project is to analyze Hangzhou DAC Biotech's research resources to recommend optimizations to their current operations. We understand the extent of Hangzhou DAC Biotech's needs by identifying the methods they use to obtain research literature. We perform a comparative analysis of different types of literature resources at their disposal. A series of interviews and surveys help us understand the key features that are most valuable to the researchers. After evaluating their needs, we recommend changes to better serve the company.

2. Background (Literature Review)

Cancer is one of the major non-transmittable diseases responsible for millions of deaths worldwide each year (Chen, 2016). Research efforts have grown exponentially over the years to combat this disease. Subsequently, the ability to effectively store, retrieve, and analyze information and research literature on this subject has made the use of literature databases significant. In this chapter, we provide an overview of literature databases and various resource sharing options. We identify the different features involved in determining the effectiveness of various literature databases. We address the different options for open access or subscription-based literature databases, along with the options and challenges of sharing literature databases among institutions.

2.1 Literature Databases

Databases are an integral part of modern day research and business, providing an efficient way to store, retrieve, and analyze massive amounts of data. These databases allow information to be added, removed, or changed quickly and efficiently (Garcia-Molina, Ullman, & Widom, 2002). A research literature database, also known as a bibliographic database, is a large collection of text-based information such as books, abstracts, and scholarly journals (Trawick and McIntyre, 2003). Over the past 200 years, the number of scientific journals and articles published has been increasing steadily. By the end of 2014, there were 28,100 peer-reviewed journals with 2.5 million articles published yearly (Rallison, 2015). Many of these articles are available through online literature databases. We discuss some of these popular research literature database resources in Section 2.3.

Most research papers lay behind steep paywalls set by publishers. These paywalls hinder the literature access of many small and medium sized companies with budgetary restrictions. To stay up to date, biotechnology researchers need access to many articles from a wide variety of journals, ranging from biology to toxicology (Lyman, 2011). Therefore, they need to purchase thousands of dollars of subscriptions or pay 30-50 USD per article, which is not feasible on a small start-up budget (Zhou, X.M., personal communication, November 11, 2016). The cost of academic research papers not only affects small biotechnology companies but also most researchers. Professor Michael Eisen at the University of California, Berkeley notes that because

of the pricing of scientific journals, the only scientists who have full access to the large amount of publications are those at extremely well-funded Western universities (Murphy, 2016).

Steep subscription prices to the main scientific journals sparked many scientists to support open access initiatives. Open access is a movement that pushes for scientific literature to be open and accessible by all. In 2015, the United States Congress passed legislation requiring tax-funded research be made publicly available 12 months after publication (Fair Access to Science & Technology Research Act (FASTR) FAQ, 2016).

However, despite the trend in open access journals, many researchers still do not have access to all the literature they need (Lyman, 2011). For this reason, researchers around the world are turning towards more controversial sites that claim to be dedicated to open access. Sci-Hub is a well-known and widely used controversial site. Section 2.3 further discusses Sci-Hub and its controversy along with two reputed open access literature resources.

2.2 Important Factors for Literature Resources

Five factors are important in describing literature database options for a user. These factors include notification systems, number of articles, cost, search options, and relevance of search results. Along with these five factors, we consider whether or not access to this resource in China requires VPN.

2.2.1 Notification systems

As cancer research efforts advance, researchers need to stay up to date with the developments in this field. Many literature resources offer a built-in notification system allowing users to set alerts that notify them when new publications pertaining to a preset search become available. They deliver these alerts through periodic emails or through a Rich Site Summary (RSS) feed. An RSS feed is a document that contains standardized data designed to be read by other websites or software (LibGuides: RSS, email, & table of contents alerts: Intro: What is RSS?, n.d.).

2.2.2 Number of articles

The number of articles a literature resource provides is an important factor when considering different platforms. The number of articles indexed reflects its scope and value. Being able to search through more articles at once is beneficial and saves time while researching.

2.2.3 Cost

Cost is important to consider, especially for smaller-scale businesses like Hangzhou DAC Biotech. Smaller companies are unlikely to have the financial resources to purchase the most advanced platforms available. The cost of different subscriptions depends on the number of users, type of institution, and amount of information accessed (Neumann, personal communication, October 7, 2016). Despite the increase in the number of literature resource platforms, many are not easily accessible. While some are completely free or partially free (only the abstract section is free), most require a paid subscription to access full documents.

2.2.4 Advanced search options

Most literature resources offer advanced search options. The advanced search options include searching using Boolean operators and the ability to apply additional criteria to any search. Boolean operators allow the user to search using terms such as AND or OR. These operators can either narrow or broaden a search. Ninety percent of users do not utilize logical connectors or query operators, even though they tend to provide more relevant search results (Eastman & Bernard, 2003). Another advanced search option is the ability to set filters to limit types of search results. Although most literature resources have similar advanced search options, the filters offered can vary depending on the resource (Kelvin Smith Library, 2016). Many literature resources also allow the results of a query to be sorted in various ways, such as by relevance or publication date.

2.2.5 Relevance of literature database search results

Assessing the effectiveness of information retrieval in literature resource platforms requires an understanding of the word “relevance.” This is a subjective term, so two users with the same question, or query, may judge the relevance of the same document differently; relevance is dynamic and depends upon context. Relevance is multifaceted because information has to relate to the query while being credible, specific, useful, clear, etc. An effective information retrieval system considers all of these qualities to present the most relevant information to the user (Ceri, 2013).

Relevance is important for evaluating the effectiveness of a literature resource. The book Web Information Retrieval mentions that the main objective of any information retrieval system is user satisfaction (Ceri, 2013). Therefore, someone evaluating an information retrieval system must consider the user and the user’s behavior. Because of this, many scientists find human judgment is the best way to define relevance of search results (Bar-Ilan, Mat-Hassan, & Levene, 2006). Studies using this method ask users to rank a certain number of results retrieved based upon which they think are most relevant to the query (Ceri, 2013).

2.2.6 Virtual private network

A virtual private network (VPN) is a tool for accessing another private network over the internet. By using a VPN, a user can access IP addresses and websites which are blocked in a certain region. It does this by routing all internet traffic through the VPN server, meaning the connection attempts to specific websites are made by the VPN server, in a different location. Thus a user with a Chinese IP address can use a VPN to access many websites that would otherwise be blocked.

2.3 Online Literature Database and Literature Resources

With the increase in number of online resources, researchers have many options for accessing online literature resources. This section introduces the subscription-based literature resources ScienceDirect and DeepDyve. These literature resources grant the user access to full-length articles for a yearly or monthly fee. NCBI and Google Scholar grant the user free access to abstracts. As an outcome of the open access movement, the Public Library of Science (PLOS) and the Directory of Open Access Journals (DOAJ) are two literature resources that require no subscription costs and all articles are full-length and free to access. We also briefly discusses another outcome of the open access movement, Sci-Hub, and why it is not a credible source of literature in the research community. For a summary of the characteristics of the each literature resources see Table 1.

2.3.1 ScienceDirect

ScienceDirect is an online literature database managed by the publisher Elsevier. It houses over 3,800 journals and 35,000 books (Elsevier, 2016). In total, ScienceDirect has over 13 million articles (ScienceDirect, 2016). Users have many search options; they can use Boolean operators and can filter by subject and date (Elsevier, 2016). Guest users can view abstracts for free and set up alerts, although viewing many of the full-length articles requires a subscription (Elsevier, 2016). A subscription for a small company with 30 employees costs 42,360 USD per year, with an 8% increase after the first year (ScienceDirect sales rep., personal communication, December 6, 2016). Outside of subscription articles, ScienceDirect has 250,000 open access articles available for guest users to view for no cost (ScienceDirect, 2016).

To stay up to date with recently published literature, ScienceDirect offers a built-in notification system. ScienceDirect's system allows the user to rename the saved search. It also allows the user to denote a frequency of either weekly, daily, or monthly for the search to be repeated in order to obtain recent articles. A user does not need VPN to access ScienceDirect in China.

2.3.2 DeepDyve

DeepDyve is an online literature database that allows users to search and view over 12 million full-length articles from multiple publishers (DeepDyve, 2016). DeepDyve's mission is "to empower information professionals worldwide by making authoritative research more simple and affordable to access" (DeepDyve, 2016). Users receive a 20% discount when purchasing articles from publishers through DeepDyve. It offers individual yearly subscriptions at 360 USD per year. It also offers group rate discounts with larger groups getting greater discounts. For 30 users, a group subscription to DeepDyve costs 10,200 USD (DeepDyve sales rep., personal communication, November 9, 2016).

DeepDyve offers advanced search features such as filters for date, author and journal. DeepDyve also offers plugins for Pubmed and Google Scholar linking to DeepDyve's full-text. One can easily create alerts for searches through email, but it is not customizable. It automatically emails the user suggestions for additional articles to read based on what the user has read. Accessing Deepdyve in China does not require a VPN.

2.3.3 NCBI: PubMed

The National Center for Biotechnology Information (NCBI) offers the literature resource PubMed, a free tool many researchers use for online medical research (Lu, 2011). It contains over 26 million citations (NCBI, 2016). NCBI is a United States government-funded resource that provides a large collection of online resources for biological information and data (NCBI, 2015). The advanced search options have 41 different criteria to filter searches, ranging from grant number to language to editor name. NCBI provides a notification system for PubMed. Its notification system allows users to change frequency, format, and number of items sent in the email alerts. A user does not need VPN to access NCBI in China.

2.3.4 Google Scholar

Google Scholar is a search tool used to search academic literature. It obtains data from academic publishers, professional societies, online repositories, universities and other websites. It allows its users to search across many disciplines and sources, directing users to databases matching their search (Jain & Raut, 2011). A 2014 study estimates Google Scholar searches through about 160 million documents, though the methods for determining this estimate are inconsistent (Orduna-Malea, Ayllón, Martín-Martín, & López-Cózar, 2014).

In addition to standard Boolean operators, users can search in particular journals, search by a publication date range, or search by author. Creating a search alert on Google Scholar is also simple and requires no account. The ease of use is beneficial, but there are no options to customize the frequency of the notification emails; the user can only choose the search query and number of alerts. Searching on Google Scholar is free, but not all articles indexed have full-text available. To access these articles, the user must purchase them from the publisher. Though Google Scholar is one of the most used search tools in the world, a user can only access it through VPN in China.

2.3.5 Directory of Open Access Journals

The Directory of Open Access Journals (DOAJ) is an online literature database that pulls from 9,000 open access journals, housing over two million articles with subjects ranging from science and technology to humanities (Directory of Open Access Journals, 2016). Its advanced search features allow for filtering the search term by title, publisher information, or subject. Sorting options for those results consist of date added to the database, publication date, relevance or title. There is no option for a saved search alert, although one can use an RSS feed. A user does not need VPN to access this literature database in China.

2.3.6 Public Library of Science

The Public Library of Science (PLOS) is an open access scientific literature database that contains eight journals with the goal of transforming communication in the research world. The over 160,000 peer-reviewed articles this literature database offers are all free to users (Public Library of Science, n.d.). Its advanced search features allow the use of Boolean operators and many filters such as title, abstract, date published, issue number, and author. The results can then be sorted by either relevance or date published. The option to save a search allows the user to rename the search and choose between weekly or monthly updates. A user must use a VPN to search this literature database in China.

2.3.7 Sci-Hub

Sci-Hub is a controversial online resource that downloads scholarly articles off of subscription databases. Its methods of obtaining these articles are controversial and some claim it to be illegal (Murphy, 2016). Sci-Hub hosts close to 50 million articles, larger than most legal resources (Mcnutt, 2016). Regardless of its controversy, scientists around the world continue to turn to Sci-Hub to access research articles (Bohannon, 2016). Some scientists in developing countries feel sites such as Sci-Hub are their only options to obtain the large amounts of material needed to conduct research (Bohannon, 2016). Despite its growing use, it remains a controversial avenue in the open access and scientific publishing field, as many open access proponents do not support Sci-Hub's illicit means of gaining access to articles (Murphy, 2016).

Table 1. Literature resource summary

	ScienceDirect	DeepDyve	NCBI: PubMed	Google Scholar	DOAJ	PLOS
Notification System	Yes	Yes	Yes	Yes	No	Yes
Number of Articles (in millions)	13	12	26	160	2	.16
Type of Payment	Subscription	Subscription	Pay Per Article	Pay Per Article	Free	Free
Search Options	Yes	Yes	Yes	Yes	Yes	Yes
Needs VPN	No	No	No	Yes	No	Yes

2.4 Literature Resource Sharing Options

With the constant rise of subscription prices, there is a significant need for alternative resource sharing options. According to EBSCO Information Services, "overall effective publisher price increases are expected to be in the range of 4 to 6 percent in 2016." (EBSCO Releases Serial Price Projection for 2016, 2015). Some options to mitigate these increasing costs include forming or joining consortia and resource collaboration.

2.4.1 Consortia

The high cost of subscriptions, especially for smaller institutions, has led to the development of consortia. A consortium is an association, typically of several business companies (Consortium, n.d.). In some cases, larger business consortia and companies will acquire smaller biotechnology companies and integrate them into the larger company through business deals.

Some of the key areas to consider when forming a consortium include developing a mission, determining scope, and targeting potential members (Updegrove, 2013). Creating a consortium requires a time investment, as the planning stages require careful deliberation and consideration. A concise mission and scope must be developed, along with a detailed description

of all aspects involved in achieving that mission (Updegrove, 2013). Once a consortium is established it requires little maintenance, a small budget, and few staff in order to function.

Consortia offer members long-term benefits. With regard to literature subscriptions, members gain greater bargaining power with publishing companies and generally gain access to 10-15% more information (Kaygusuz, 2008). In addition, other companies participating in the consortium have a potential for cooperation in other areas besides research literature access. The largest cost is during the establishing process of the consortium, and there is no guarantee the consortium will succeed.

In some cases larger companies or consortia will seek out smaller innovative companies and form partnerships with them. Ambrx is an example of a small biotechnology firm that develops Antibody Drug Conjugates benefiting from partnership with a large corporation. Its most advanced ADC, is currently being tested in its first-in-human study. They also just finished a 45 million USD round of financing in August (Ambrx Inc., 2016). Much of their success can be attributed to being acquired by a Shanghai consortium in 2015 (Ambrx Inc., 2016).

2.4.2 Resource collaboration

Many biotechnology companies partner with large pharmaceutical institutions to help develop products in their pipeline. When striking a deal, the biotechnology partner can request access to the pharmaceutical company partner's literature resource subscriptions as a part of the agreement. For a large pharmaceutical firm with thousands of staff members, the cost to add access for 20-30 scientists from a small biotechnology company is small. Additionally, access to literature facilitates the ability of the biotechnology partner to advance joint and personal projects (Lyman, 2011).

3. Methodology

The goal of this project is to analyze Hangzhou DAC Biotech's research resources to recommend optimizations to their current operations. To meet this goal, we have three principal objectives:

1. Evaluate current research methods and resources
2. Investigate alternative research resources
3. Analyze literature resources

These three objectives require interviews and/or surveys to gather the necessary information. Interviews with librarians at WPI and Hangzhou Dianzi University (HDU), as well as online research, help us discover the availability of different resources. We interview and survey employees to understand the challenges Hangzhou DAC Biotech's researchers face when conducting online research. This chapter discusses the methods we design to accomplish these objectives.

Language barriers and cultural differences are a challenge we face with our interviews and surveys; they lead to miscommunication between the interviewer and interviewee. A Chinese translator is present to mitigate miscommunication during interviews. As setting up one-on-one interviews requires more time, we use surveys to collect information company-wide. We also include Chinese translations on our employee survey.

3.1 Evaluate Current Research Methods and Resources

To understand the current methods and resources in use at Hangzhou DAC Biotech, we interview and survey the employees. These interviews and surveys include questions to determine the range of literature resource platforms frequently used. These questions help us understand the accessibility of resources and determine the types of resources they value. We gauge the challenges employees face to obtain literature and their causes. We inquire about the different methods researchers employ to try to solve these challenges. The surveys also help us determine if the problems are company-wide.

We interview the sponsor to identify the resources the company provides as well as their costs. Through interviews and surveys of employees we gauge whether the researchers utilize the

available resources. The sponsor interview can be found in Appendix I. The employee survey and interview can be found in Appendices B and F.

3.2 Investigate Alternative Research Resources

We explore notification systems, collaborations, and literature resources as potential solutions to some of the challenges researchers face accessing research literature. We investigate notification systems to help researchers stay up to date with publications. We analyze potential collaborations as an alternative method to gain access to full-length articles. This section discusses the notification system and user guide we implement as well as the potential collaborations we explore. Section 3.3 discusses the literature resources we evaluate.

3.2.1 Notification system and guide

We investigate the notification systems of the most used resources. Upon our sponsor's request, we also examine the notification system of the United States Patent and Trademark Office. We distribute a survey to gauge the kind of notification system each employee wants. As some notification systems require accounts, we create accounts for the employees if necessary. An email is sent to all the employees explaining the activation process. We create a test alert on each new account using a common keyword so employees can see the format of the notifications. To ensure that there is no confusion, we create a guide with directions on how to manage the account, perform advanced searches, and add/change/remove alerts. We send the guide and account information to all the users to allow them to edit the notification. Survey questions can be found in Appendix C.

3.2.2 Potential collaborations

Understanding the relationships between Hangzhou DAC Biotech and other research institutions, companies, and universities lets us better comprehend the external resources available. By interviewing our sponsor about the company's current relationships, potentially sharing literature resources, and working with other institutions in the future, we gain an understanding of the feasibility of potential collaborations. This interview includes questions regarding the company's plans and collaborations to help recommend systems for the present and future. The sponsor interview is found in Appendix I.

We examine the possibility of Hangzhou DAC Biotech forming or joining a consortium. As WPI is a part of consortium, we gain insight through an interview with the WPI librarian on the terms and benefits of membership. This allows us to examine if a similar system will work for Hangzhou DAC Biotech. The WPI librarian interview is found in Appendix H.

In addition to collaborations in the form of consortia, at the request of our sponsor, we interview library officers to inquire whether a collaboration is plausible. We interview HDU library officers to analyze the feasibility of HDU library sharing resources with Hangzhou DAC Biotech. The HDU librarian interview is found in Appendix G.

3.3 Analyze Literature Resources

We compare literature resources to suggest a resource that can provide better access to full-length articles. We use a decision matrix to analyze the resources based on predetermined characteristics such as relevance of search results. This section discusses methods for determining weights for the characteristics and conducting the relevance test.

3.3.1 Literature resource comparison

We create a weighted decision matrix using the factors outlined in our background. The matrix helps determine the usefulness of different literature resources for Hangzhou DAC Biotech. We conduct an interview with our sponsor to understand what features are most relevant to Hangzhou DAC Biotech. This interview determines the weights for the decision matrix. Qualities our sponsor deems more important are weighted more heavily.

3.3.2 Relevance of search results

Calculating resource search relevance is a more involved process than determining the other factors of each resource and thus takes its own methods. To determine search result relevance of each literature resource platform, we identify what the researchers at Hangzhou DAC Biotech find to be the most relevant. We distribute electronic surveys through email to all Hangzhou DAC Biotech researchers. The survey asks for specific keywords relevant to their research fields, similar to terms they use while performing research. We use the keywords this survey provides to search the literature resource platforms we compare. As Hangzhou DAC Biotech consists of three departments that each focus on different aspects of ADC research, we

collect input from all three departments. Thus, there are specific surveys for each department based on the keywords researchers provide. This ensures each researcher is familiar with the keywords they are evaluating. We place screen captures of these searches into an online survey that we send to the researchers via email. The survey asks the researchers to compare the first five search results from each literature resource and rate each result as relevant, somewhat relevant, or not relevant. The specific keyword survey and relevance survey can be found in Appendices D and E.

4. Results and Analysis

This chapter presents our team's research findings and an analysis of these findings to improve research at Hangzhou DAC Biotech. We analyze Hangzhou DAC Biotech's current research methods and literature resources to identify any problems or areas for improvement. We examine literature database alternatives, potential collaborations with other institutions, and notifications systems.

4.1 Analysis of Current Methods and Resources

Surveys and interviews with Hangzhou DAC Biotech researchers allow us to identify the challenges they encounter and the resources they frequently use. In this section, we evaluate the different resources used and the problems associated with them.

4.1.1 Finding: Hangzhou DAC Biotech primarily utilizes NCBI and Google

More than 50% of the company uses NCBI and Google (Figure 1). NCBI is a popular option for the researchers because it is free to use and specifically a biotechnology resource. Google is also a popular option because of its comprehensive research coverage, despite requiring VPN to use in China. Though these two platforms are the most popular, researchers utilize other options while searching for articles online. The employee survey results are found in Appendix J.

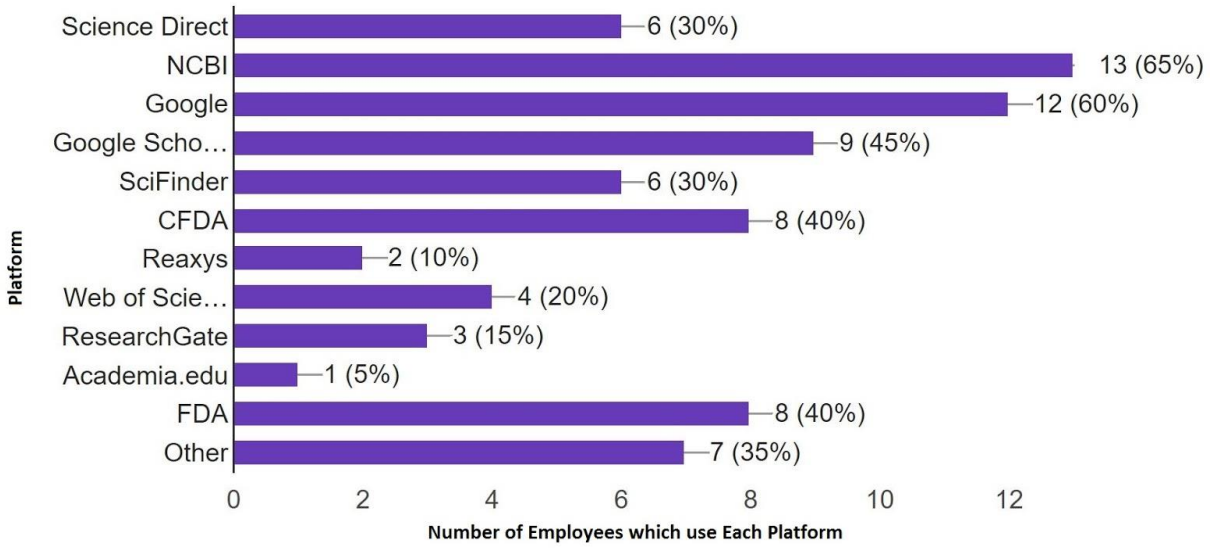


Figure 1. Online resources used

4.1.2 Finding: Hangzhou DAC Biotech's ScienceDirect subscription does not provide value

Hangzhou DAC Biotech has only one subscription to ScienceDirect. They currently spend about 400 USD per year on ScienceDirect, which gives them access to several journals. Only 30% of researchers surveyed utilize ScienceDirect. Our sponsor revealed that even with this small yearly subscription, the company still pays per article for most Elsevier publications. Therefore, this subscription does not appear to benefit the company. The sponsor interview and employee survey results can be found in Appendices I and J.

4.1.3 Finding: Researchers at Hangzhou DAC Biotech have trouble keeping up to date with new publications

Nearly half of the respondents have trouble keeping up to date with newly published research. Most of the researchers look for new research once or twice a week. However, interviews with our sponsor suggest that the researchers occasionally forget to look for new publications. Despite having difficulty staying up to date, none of the survey respondents utilize online notification systems. The sponsor interview and employee survey results can be found in Appendices I and J.

4.1.4 Finding: Researchers at Hangzhou DAC Biotech can not access all the full-length articles they need

Almost 90% of survey respondents have trouble accessing full-length research articles. Our sponsor estimated that they access approximately half of the articles they need through open access journals or China's Foreign Publication Database. Because the current ScienceDirect subscription does not cover all material needed, they purchase paywalled articles for 30 USD or more. Some researchers occasionally spend their own money to pay for full-length research papers. Many contact friends or former colleagues associated with institutions that have subscriptions with access to full-length articles. The underlying problems for researchers are the high cost of research articles and Hangzhou DAC Biotech's limited budget for literature resources. The company needs a more cost-effective way to access full-length papers. The employee survey results can be found in Appendix J.

4.2 Evaluation of Potential Solutions

Based on our findings, we identify and explore potential solutions to Hangzhou DAC Biotech's literature research challenges: staying up-to-date and accessing full-length articles. This section analyzes potential solutions through different notification systems and resource sharing options.

4.2.1 Finding: A notification system is an easy and effective way for researchers to stay up to date

Our sponsor expressed interest in setting up a notification system for both research articles and patents to help researchers stay up to date. When researching options, we found that most online literature databases offer simple, built-in notification systems. Some literature resources offer RSS feeds, but these require an RSS reader to use. RSS readers create inconsistency with notifications throughout a company as researchers each use their own reader. We focus on email-based notification systems because they are easy to set up and the notifications arrive as a convenient email message.

4.2.2 Finding: Licensing agreements and underdeveloped partnerships prevent Hangzhou DAC Biotech from accessing literature resources through other institutions

To address Hangzhou DAC Biotech’s lack of research literature resources, we explore sharing resources with other institutions. Upon our sponsor’s request, we contacted HDU to discover if Hangzhou DAC Biotech can collaborate with HDU’s library. During an interview with HDU library officers, we learned that HDU has a VPN that allows access to their library resources from off campus. Licensing agreements with publishing companies prevent Hangzhou DAC Biotech from utilizing these resources. The HDU library officers’ interview is found in Appendix G.

We also explore possible collaborations with institutions currently partnering with Hangzhou DAC Biotech. The company currently works with Lizhu Pharmaceuticals and the University of Science and Technology of China on experimental research projects. According to our sponsor, the partnerships with these institutions are relatively new and only between certain departments. Our sponsor stated it will be challenging for Hangzhou DAC Biotech to make new arrangements to their deals with these institutions. The sponsor interview is found in Appendix I.

4.2.3 Finding: Creating a consortium requires significant time and monetary investment

Another possible option we explore is consortia. Developing a consortium from the ground up requires more time and effort than joining an already-established consortium. During interviews with our sponsor, we learn that the company has limited resources and all efforts currently focus on moving their ADCs into clinical trials within the next three years. Once the company's budget increases, this may become a viable option. If Hangzhou DAC Biotech joins a pre-existing consortium it may provide improvement for research literature access. Few commercial consortia exist in China, but if the company finds a suitable consortium to join it can give the company access to the literature resources they need. The sponsor interview is found in Appendix I.

4.3 Literature Resource Comparison

In this section, we compare the six literature resources using the factors outlined in our background: cost, relevance of search results, advanced search options, need for VPN, number of articles, and notification systems. We develop a weighted decision matrix containing all the factors and literature resources to determine the best option for Hangzhou DAC Biotech. We use a skewed weighting system which places emphasis on the most desired factors. Our sponsor rated the importance of each factor as either very important, somewhat important, or not important. These factors receive a weighting of 10, 3, or 1 respectively. The most important factors are cost, need for VPN, and search result relevance. We calculate each factor individually utilizing different criteria and scale all results between 0 and 5. Detailed calculations can be found in Appendix O.

4.3.1 Finding: ScienceDirect has the most expensive subscription of the compared resources

To compare the cost of each literature resource, we evaluate them based on yearly article usage. Subscription-based literature resource platforms have a flat yearly subscription cost which provides access to all of its articles. To estimate the cost of literature resources such as NCBI and Google for our decision matrix, we estimate the cost based on amount of articles purchased per year at publisher prices. From interviews with our sponsor, we estimate the average cost of an article to be 40 USD. The estimated cost of NCBI and Google assume that each researcher at the company requires one paywalled article per month. At 360 articles per year, paywalled articles on NCBI and Google cost 14,400 USD. The sponsor interview is found in Appendix I.

Figure 2 shows the comparison of literature resource platform costs dependent on the number of articles purchased per year. Subscription and open access literature resources have a constant cost regardless of number of articles. The horizontal lines on the graph denote these literature resources. The cost of articles for Google and NCBI increases linearly depending on the number of articles researchers purchase per year. The points where NCBI and Google intersect the horizontal lines are when purchasing a subscription becomes more cost-effective than purchasing articles individually.

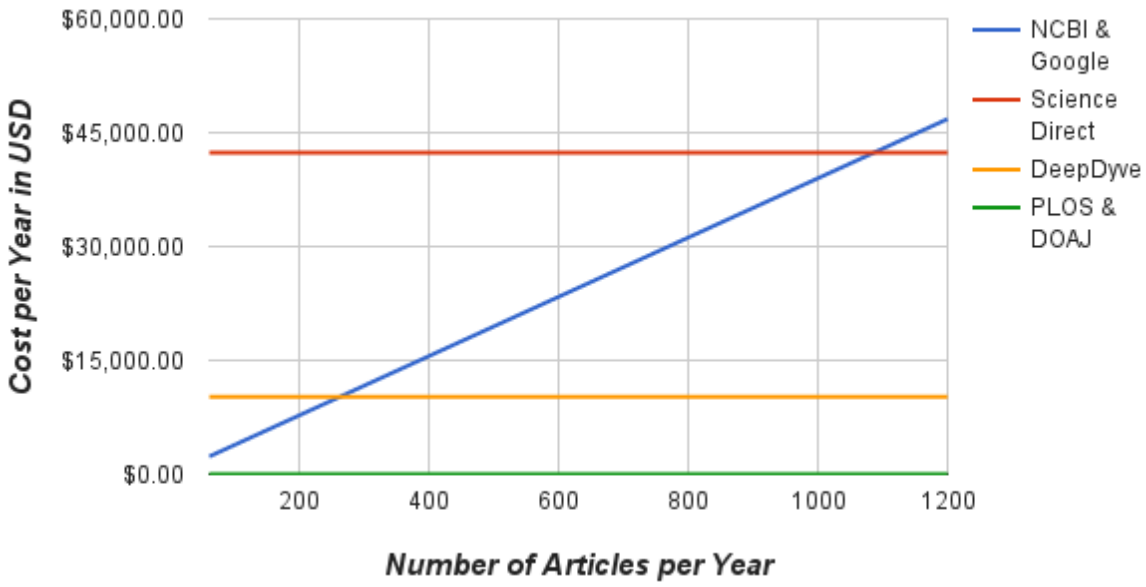


Figure 2. Cost of articles per year by number of articles

4.3.2 Finding: Google and DeepDyve provide the most relevant search results

We calculate relevance scores for each literature resource based on the responses researchers at Hangzhou DAC Biotech provide (Table 2). There is an unequal distribution of keywords for each department from our keyword survey. To compare between departments, we average the relevance of each keyword for each database evaluated by employees based on their department. These averages receive a weight dependent on how many keywords the department evaluates. Even though the chemistry department had more respondents than pharmacology, pharmacology receives a heavier weight because we evaluate more keywords for pharmacology. The specific keyword survey results are found in Appendix L.

Google and DeepDyve provide the most relevant results for the researchers' specific keywords. These results varied slightly depending on department. For example, both the biology and chemistry results show that Google provides the most relevant results, while the pharmacology results show DeepDyve has the most relevant results. Additionally, the pharmacology relevance scores are all relatively higher when compared to biology or chemistry, because one of the respondents appeared to find a majority of search results relevant. This researcher likely had loose criteria when choosing whether a result was relevant or not. A single outlier such as this does not significantly affect the results of testing.

Table 2. Relevance testing results

	Weight	PLOS	DOAJ	DeepDyve	ScienceDirect	NCBI	Google
Biology:	0.514	0.539	0.183	0.570	0.519	0.488	0.620
Chemistry:	0.189	0.364	0.284	0.530	0.391	0.370	0.574
Pharmacology:	0.297	0.823	0.499	0.900	0.874	0.785	0.813
	Weighted Average:	0.590	0.296	0.660	0.600	0.554	0.669
	Scaled to 5:	**2.95**	**1.48**	**3.30**	**3.00**	**2.77**	**3.34**

4.3.3 Finding: NCBI has the most advanced search options

We evaluate advanced search options based on three criteria: use of Boolean operators, number of filters, and ways to sort results (Table 3). If a literature resource supports operators in search queries it receives a 1. If it lacks the ability to understand these operators it receives a 0. We compare the number of filters each literature resource platform can utilize while searching. The literature resource platform with the most filters receives a 1 while we scale the others based

on their relative number of filters. We utilize a similar process to determine the scoring for result sorting. A literature resource which receives a 1 for all criteria scores a 5, and the others scale accordingly.

NCBI receives the highest score in this factor with 4.79. This is because it has the most filters out of any platforms we evaluate. PLOS is the second best, receiving a score of 4.32. While PLOS offers the largest number of ways to sort search results, it only has a little over half the number of filters NCBI has. Despite returning some of the most relevant results, DeepDyve has the least advanced search options. It is also a special case, receiving a 0.5 for Boolean operators. Most search engines automatically combine multiple keywords with OR to broaden a search. During testing, DeepDyve appeared to only recognize some operators like NOT and AND, but not OR. Appendix O has detailed calculations.

Table 3. Evaluation of advance search options

	NCBI	ScienceDirect	Google	DeepDyve	PLOS	DOAJ
Factor	Score	Score	Score	Score	Score	Score
Boolean operators	1.00	1.00	1.00	0.50	1.00	1.00
Number of filters	1.00	0.32	0.18	0.09	0.59	0.30
Number of ways to sort results	0.88	0.63	0.25	0.38	1.00	0.50
Total Weighted Score:	2.88	1.94	1.43	0.97	2.59	1.80
Scaled to 5:	**4.79**	**3.24**	**2.39**	**1.61**	**4.32**	**2.99**

4.3.4 Finding: Google Scholar indexes the most articles followed by NCBI

We research the number of articles each resource contains and compare them. The largest resource receives a score of 5 and we scale the others accordingly. Google Scholar indexes the largest number of articles when searching, but the exact number is unknown. A statistical study estimates the size of Google Scholar to be 160 million articles. Even if the true value were only 50% of this estimate, it is still several times larger than the second largest resource we explore. NCBI is the second largest platform, and is a beneficial tool for biotechnology researchers because it searches across many publishers at once. ScienceDirect, while having a strong search result relevance score, only will return articles published by Elsevier. DeepDyve has less total articles when compared to ScienceDirect, but has the advantage of searching across journals from multiple publishers. A comparison of number of articles indexed by each literature database is shown in Figure 3.

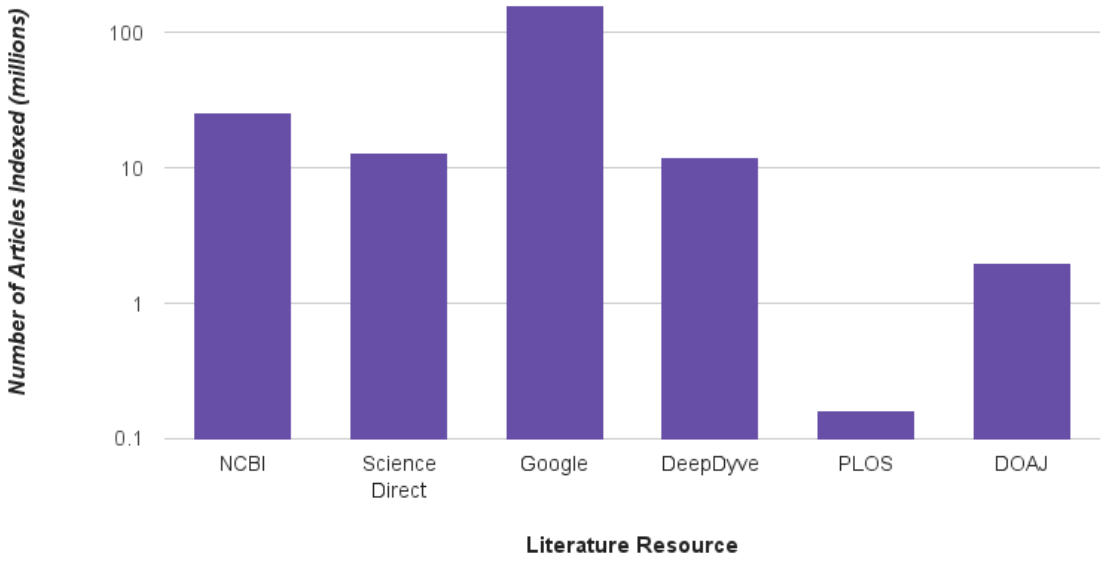


Figure 3. Number of articles indexed in millions (log scale)

4.3.5 Finding: NCBI has the most advanced notification system

We evaluate the notification systems of literature resources based on several criteria: general frequency, detailed frequency, choice of notification format, choice of number of articles sent per alert, ease of use, and ability to edit saved searches (Table 4). We select these criteria based on examinations of the features of each literature resource's notification system. General frequency means having weekly or monthly alert options, which exists for NCBI, ScienceDirect, and PLOS. Detailed frequency is a criteria unique to NCBI, allowing users the ability to choose a particular day of the week to receive alerts. All the literature resource platforms have a button on their user interfaces for quickly making alerts, except for DOAJ. DOAJ is the only platform we explore which completely lacks an email notification system. These results do not significantly impact our final recommendation because notification system features are less important when compared to search relevance and cost.

Our sponsor also specifically requests notifications from the United States Patent and Trademark Office (USPTO). Google Scholar notifications provide patent publications alerts in addition to literature publications. We also explore the Patent Application Alert Service, which is

a third party service contracted by the USPTO. This service has more customizations for searches than Google's email alerts, and allows for editing of notifications after they are set.

Table 4. Evaluation of notification systems

	NCBI	Science Direct	Google	DeepDyve	PLOS	DOAJ
Factor	Score	Score	Score	Score	Score	Score
General Frequency (weekly/monthly)	1	1	0	0	1	0
Detailed Frequency (choose day of the week)	1	0	0	0	0	0
Ability to choose format of notifications sent (abstracts/titles/etc)	1	0	0	0	0	0
Ability to choose number of articles sent per alert	1	0	1	0	0	0
Ability to easily save a search (has "Save Search" button)	1	1	1	1	1	0
Ability to edit the saved search	1	1	0	0	1	0
Total Score:	6	3	2	1	3	0
Scaled to 5:	**5**	**2.5**	**1.67**	**0.83**	**2.5**	**0**

4.3.6 Finding: The overall best literature databases are NCBI, DeepDyve, and DOAJ

Combining the scores of each factor and their weights, we generate an overall score for each literature resource (Table 5). DeepDyve is a more effective subscription-based literature resource compared to ScienceDirect. DOAJ is a superior open access literature resource compared to PLOS. NCBI is the highest rated literature resource overall. Google and PLOS receive weaker scores primarily because they require VPN to access. DeepDyve, despite having limited advanced search features, produced some of the most relevant search results at about a quarter of the cost of ScienceDirect. Even though DOAJ scored somewhat highly, it has significant drawbacks. DOAJ had the worst relevance score, and has the second fewest number of articles available. However, it is a free resource and therefore can provide benefit towards accessing full-length articles for Hangzhou DAC Biotech. If VPN were not a limiting factor, Google is one of the highest rated literature resources and is still worth consideration.

Table 5. Final decision matrix

		Science Direct		DeepDyve	
Factor	**Scaled Weights**	Score	Weighted Score	Score	Weighted Score
Cost	0.27	0.00	0.00	3.80	1.03
Search Result Relevance	0.27	3.00	0.81	3.30	0.89
Advanced Search Options	0.08	3.24	0.26	1.61	0.13
Need for VPN	0.27	5.00	1.35	5.00	1.35
Number of Articles	0.08	0.40	0.03	0.37	0.03
Notification System	0.03	2.50	0.07	0.83	0.02
Total Weights:	1.00	Total Weighted Score:	**2.52**	Total Weighted Score:	**3.45**
		DOAJ		**PLOS**	
Factor	**Scaled Weights**	Score	Weighted Score	Score	Weighted Score
Cost	0.27	5.00	1.35	5.00	1.35
Search Result Relevance	0.27	1.48	0.40	2.95	0.80
Advanced Search Options	0.08	2.99	0.24	4.32	0.35
Need for VPN	0.27	5.00	1.35	0.00	0.00
Number of Articles	0.08	0.06	0.00	0.00	0.00
Notification System	0.03	0.00	0.00	2.50	0.07
Total Weights:	1.00	Total Weighted Score:	**3.35**	Total Weighted Score:	**2.57**
		NCBI		**Google**	
Factor	**Scaled Weights**	Score	Weighted Score	Score	Weighted Score
Cost	0.27	3.30	0.89	3.30	0.89
Search Result Relevance	0.27	2.77	0.75	3.34	0.90
Advanced Search Options	0.08	4.79	0.39	2.39	0.19
Need for VPN	0.27	5.00	1.35	0.00	0.00
Number of Articles	0.08	0.80	0.06	5.00	0.41
Notification System	0.03	5.00	0.14	1.67	0.05
Total Weights:	1.00	Total Weighted Score:	**3.58**	Total Weighted Score:	**2.44**

5. Conclusions and Recommendations

This chapter presents our conclusions and recommendations as a result of our findings. We provide recommendations to Hangzhou DAC Biotech to improve their research literature access and to help researchers stay up to date with new publications. Our recommendations consider their current resources and needs.

We conclude which notification system helps Hangzhou DAC Biotech researchers most. Our findings on consortia and resource sharing allow us to formulate conclusions on the feasibility of these options for the company. We recommend the use of the NCBI notification system, a combination of literature resources and future reevaluations of collaborations. We present which combination of resources best fulfills Hangzhou DAC Biotech's needs.

5.1 NCBI's Notification System

As it has the best notification system, we recommend Hangzhou DAC Biotech researchers make use of the NCBI accounts we created for them. In addition, we suggest the researchers annually update their search alerts to ensure the future effectiveness of the notification system. By making use of the guide we provide, researchers are able to update their search alerts to reflect the current focus of their research. We suggest researchers also customize their searches to receive alerts on specific literature.

We also provide a guide to set up notifications for the United States Patent and Trademark Office's (USPTO) Patent Application Alert Service. We recommend researchers create accounts and use this service to stay up to date with patent applications. Although some researchers requested notifications from Google Scholar, its need for VPN makes it difficult for researchers to maintain in the future. Therefore we do not recommend this system. If the regulations for accessing Google Scholar in China change, then DAC should reconsider and reevaluate the use of this resource. Appendix N shows the full tutorial for recommended notification systems.

5.2 Collaborations and Consortia

When considering the company's resources, creating a consortium is not a feasible option at this time. Hangzhou DAC Biotech cannot afford the time and monetary investment forming a consortium requires. Hangzhou DAC Biotech can benefit from this form of collaboration only if a pre-existing consortium is found. Once the company has the time and other resources to invest, forming a consortium is possible. We recommend that Hangzhou DAC Biotech also reevaluate the notion of joining and forming a consortium in the future. They should prioritize joining a consortium over creating one from scratch because joining requires less resources.

In looking into other forms of collaboration, resource sharing with other institutions is not a viable option at this time. Hangzhou DAC Biotech's partnership with Lizhu Pharmaceuticals and the University of Science and Technology of China are currently too new and underdeveloped to allow for literature resource sharing. If the relationships with these institutions continue to develop and they strike more agreements, we recommend Hangzhou DAC Biotech request access to their literature resources as a part of the agreement. Hangzhou DAC Biotech might request access to the literature resources of pharmaceutical companies they work with in the future.

5.3 DeepDyve Subscription

Compared to ScienceDirect, a subscription to DeepDyve is a more cost-effective option for purchasing full-length articles. This allows Hangzhou DAC Biotech researchers to access full-length articles from various publishers, including Elsevier, at prices more suitable to Hangzhou DAC Biotech's monetary limitations.

We recommend the 400 USD spent on one ScienceDirect account be redirected to one DeepDyve account, giving them a 20% discount on full-length article PDF purchases which can be shared with the entire company. The Deepdyve subscription allows a user to view the full-text of articles online, but for the entire company to effectively utilize the view-online feature, multiple accounts are needed as only person can be logged on at a time. Additionally, another option for Hangzhou DAC Biotech is to purchase one or two more accounts to share, assigning a specific person in charge of each account to purchase the articles the researchers need at the 20% discount.

In addition to switching to a DeepDyve subscription, we recommend Hangzhou DAC Biotech periodically reevaluate the relevance of their literature resource subscriptions. As the company grows and the number of employees increases, the cost and relevance comparison of the literature resources may change. The relevance of search results may also change as their research focus changes. Once the company has a larger income, we recommend they consider looking into other literature resources.

5.4 Literature Resources

As our decision matrix reflects, NCBI is a powerful tool for Hangzhou DAC Biotech researchers to retrieve specific research literature. PLOS, Google Scholar, and DOAJ are other literature resource options, however, Section 4.3.6 shows, they are less suitable to Hangzhou DAC Biotech. Using the more suitable tool, NCBI, in conjunction with other literature resources increases researchers' ability to access full-length articles. Searching ScienceDirect and DeepDyve can supplement NCBI as they have a higher number of relevant results than NCBI and users are able to search without subscriptions.

5.5 Limitations

Time restraints limit us from being able to look into all avenues, more than six literature resources could have been analyzed. Some of our statistics maybe be skewed as they come solely from our sponsor. The views of our sponsor may potentially differ from the views of the company as a whole. The cost calculations used in the decision matrix were made under the assumption that each researcher needs one article per month. This limits our project as this estimated number of articles needed per year may differ from the actual needs of the researchers. We provide the Excel sheet with all the calculations to our sponsor to allow the decision matrix to be updated with a more accurate cost analysis of NCBI and Google Scholar. Although mitigated with the presence of a translator, there was still some miscommunication in surveys and interviews. The small number of respondents to our relevance and specific keyword survey limit our relevance test data. More respondents would have allowed for a greater number of keywords to be tested.

References

Ambrx, Inc. (2016). Retrieved December 07, 2016, from http://ambrx.com/

Bar-Ilan, J., Mat-Hassan, M., & Levene, M. (2006, July 14). Methods for comparing rankings of search engine results. *Computer Networks, 50*(10), 1448-1463. doi:10.1016/j.comnet.2005.10.020

Bohannon, J. (2016, April 28). Who's downloading pirated papers? Everyone. *Science*. doi:10.1126/science.aaf5664

Bornmann, L., & Mutz, R. (2015, April 29). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology, 66*(11), 2215-2222. doi:10.1002/asi.23329

Ceri, S. (2013). *Web Information Retrieval*. Berlin: Springer.

Chen, W., Zheng, R., Baade, P. D., Zhang, S., Zeng, H., Bray, F. . . . He, J. (2016, January 25). Cancer statistics in China, 2015. *CA: A Cancer Journal for Clinicians, 66*(2), 115-132. doi:10.3322/caac.21338

Consortium [Def. 1] (n.d.) Oxford Dictionaries Online, Retrieved November 13, 2016, from https://en.oxforddictionaries.com/definition/consortium

DeepDyve, Inc. (2016). Retrieved December 07, 2016, from https://www.deepdyve.com/

Directory of Open Access Journals (2016). Retrieved from https://doaj.org/

Eastman, C. M., & Jansen, B. J. (2003, October). Coverage, relevance, and ranking. *ACM Transactions on Information Systems, 21*(4), 383-411. doi:10.1145/944012.944015

EBSCO Releases Serial Price Projection for 2016. (2015, September 29). Retrieved December 07, 2016, from https://www.ebsco.com/news-center/press-releases/ebsco-releases-serial-price-projection-for-2016

Elsevier (2016). Retrieved December 07, 2016, from https://www.elsevier.com/

Fair Access to Science & Technology Research Act (FASTR) FAQ. (n.d.). Retrieved December 07, 2016, from http://sparcopen.org/our-work/fastr/faq/

Garcia-Molina, H., Ullman, J. D., & Widom, J. (2002). *Database systems: The complete book*. Upper Saddle River, NJ: Prentice Hall.

Jain, V., & Raut, D. (2011). Medical literature search dot com. *Indian Journal of Dermatology, Venereology, and Leprology, 77*(2), 135. doi:10.4103/0378-6323.77451

Kaygusuz, A. (2008, November 20). Universities and University Libraries in Turkey. *LIBER Quarterly, 18*(3-4), 480. doi:10.18352/lq.7947

LibGuides: RSS, email, & table of contents alerts: Intro: What is RSS? (n.d.). Retrieved December 07, 2016, from http://libguides.mit.edu/rss

Lu, Z. (2011). PubMed and beyond: A survey of web tools for searching biomedical literature. *Database, 2011*(0). doi:10.1093/database/baq036

Lyman, S. (2011, July 11). Industry access to the literature. *Nature Biotechnology, 29*(7), 571-572. doi:10.1038/nbt.1909

Mcnutt, M. (2016, April 28). My love-hate of Sci-Hub. *Science, 352*(6285), 497-497. doi:10.1126/science.aaf9419

Murphy, K. (2016, March 12). Should All Research Papers Be Free? Retrieved December 07, 2016, from http://www.nytimes.com/2016/03/13/opinion/sunday/should-all-research-papers-be-free.html

National Center for Biotechnology Information, U.S. National Library of Medicine (2016). Retrieved December 07, 2016, from https://www.ncbi.nlm.nih.gov/

Orduna-Malea, E., Ayllón, J. M., Martín-Martín, A., & López-Cózar, E. D. (2015, May 31). Methods for estimating the size of Google Scholar. *Scientometrics, 104*(3), 931-949. doi:10.1007/s11192-015-1614-6

Perushek, D. E., & Douglas, A. (2014). Culture, politics and university library consortia in China and the US. *Library Management, 35*(8/9), 594-606. doi:10.1108/lm-03-2014-0039

Public Library of Science. (n.d.). Retrieved December 07, 2016, from https://www.plos.org/

Rallison, S. (2015, March). What are Journals for? *The Annals of The Royal College of Surgeons of England, 97*(2), 89-91. doi:10.1308/003588414x14055925061397

ScienceDirect (2016). Retrieved from http://www.sciencedirect.com/

Trawick BW, McEntyre J. Chap 1, Bibliographic databases. In: Sansom C, Horton R,, editors. The Internet for Molecular Biologists: A Practical Approach. Oxford: Oxford University Press; 2004. pp. 1–16.

Updegrove, A. (2013). The Essential Guide to Consortia and Standards. Retrieved November 10, 2016, from http://www.consortiuminfo.org/essentialguide/#.WChaE_l96Uk

Appendix A: Sponsor Description

Hangzhou DAC Biotech is a company with the primary goal of developing cancer-fighting antibody drug conjugates (ADC's). Formed in 2012, the company resides in the Hangzhou Economic and Technological Development Area (HEDA) near the Qiantang River. The lead developers of the ADCs are senior-level scientists and project designers. These leaders work with a research team of approximately 30 other scientists specializing in either biology, chemistry or pharmacology. They receive monetary support from the local government of HEDA. Hangzhou DAC Biotech consists of three major departments. The first department researches preclinical drug effects, metabolism, and toxicology using animal experiments. The second is dedicated towards creating an antibody that can combat cancer, with the third creating the ADC system to deliver the antibody (Hangzhou DAC Biotech, 2014). These departments consist of teams of master level or higher scientists. Through the investment of almost 40 million RMB in 2013, the company holds four patents in their field. They plan to move their ADCs to clinical trials within the next three years.

Appendix B: Employee Survey

员工调查

Thank you for taking the time to fill out our survey. We are a team of students working on our Interactive Qualifying Project. This project's goal is to help researchers at Hangzhou DAC Biotech find full-length research literature more effectively. We believe you will help us understand different aspects of the overall problem. The results of this survey will also help us analyze different possible solutions. Your participation in this survey is completely voluntary and you may skip any questions you don not want to answer. Please remember that your answers will remain anonymous. No names or identifying information will appear in any of our project reports or publications. This project is a collaborative project between WPI, Hangzhou Dianzi University and Hangzhou DAC Biotech. Your participation is highly appreciated.

十分感谢您能接受我们的采访。我们是来自美国马萨诸塞州伍斯特理工学院（WPI）的一只学生团队。这个项目旨在帮助 DAC 生物科技的研究人员能够更快地搜索文献。我们邀请您来做我们的采访对象，相信您会帮助我们从不同角度理解这个问题，并帮助我们分析不同的可能解决问题的方法。这个项目的成果可能或多或少也会惠及您，我们坚信这项研究会在之后引出更加详尽的方案来解贵公司遇到的问题。您的参与是完全自愿的，您可以随时退出并跳过任意问题。请注意您的回答是匿名的，不会有名字及身份认证信息出现在任何调查问卷，项目报告或出版刊物中。该项目由 WPI，杭州电子科技大学，DAC 生物科技联合完成。非常感谢您的参与。

Job title, 职称:_____
Department, 部门:_____
Area/s of research, 研究领域:_____
Questions:

1. **About how often do you look for new research? 您多久寻找一次新的项目？**
   - ☐ Every day (每天)
   - ☐ Every other day (每隔一天)
   - ☐ Twice a week (一周两次)
   - ☐ Once a week (一周一次)
   - ☐ Less than once a week (一周不到一次)

2. **Do you use any of the following database platforms for research? (Check all that apply) 您在研究过程中有使用以下哪种平台吗？ (可多选）**

- ☐ ScienceDirect
- ☐ NCBI
- ☐ Google
- ☐ Google Scholar
- ☐ SciFinder
- ☐ CFDA
- ☐ Reaxys
- ☐ Web of Science/Web of Knowledge
- ☐ ResearchGate
- ☐ Academia.edu
- ☐ FDA
- ☐ Other (Please specify in the box below)其他（请在方框内具体标明）

3. **Which of the above database platforms is your favorite/primary?**
在以上项您最喜欢或您认为最重要的数据库是哪个？

4. **Are there any resources you wish you had access to that you currently don not have access to? If so, please specify in the box below. 是否有任何您想要访问的数据库？请在方框内具体标明.**

5. **Do you have trouble searching for research?您在查找文献的过程中遇到过困难吗？**

- ☐ Yes 有
- ☐ No 没有

6. **If yes, how often? 如果你有，多久？**
   - ☐ Always（绝大多数情况下）
   - ☐ Most of the time（大多数情况下）
   - ☐ Sometimes（有时）
   - ☐ Rarely（很少

Why?为什么？

7. **Do you have trouble keeping up to date with recently published research? 您会保持更新最新刊登的文献吗？**
   - ☐ Yes
   - ☐ No

8. **If yes, how often?如果你有，多久？**
   - ☐ Always（绝大多数情况下）
   - ☐ Most of the time（大多数情况下）
   - ☐ Sometimes（有时）
   - ☐ Rarely（很少)

Why?为什么？

9. **Do you use any kind of notification system for new publications?您是否使用通知系统的新出版物？**

☐ Yes 使用　　☐ No 不使用

If yes, what is it? 如果使用, 它是什么？

10. **Do you have trouble accessing full-length papers while researching? 您在查找文献过程中，又遇到是否能阅读全文的权限问题吗？**

☐ Yes 有　　☐ No 没有

11. **If yes, how often?如果你有，多久？**

☐ Always（绝大多数情况下）　　☐ Sometimes（有时）

☐ Most of the time（大多数情况下）　　☐ Rarely（很少）

Why?为什么？

12. **How often do you have to purchase papers which are not covered by Hangzhou DAC Biotech's current subscriptions?**

☐ Always（绝大多数情况下）

☐ Most of the time（大多数情况下）

☐ Sometimes（有时）

☐ Rarely（很少）

☐ Never (决不)

13. **Do you use your own money to pay for full-length research papers?你使用自己的钱来支付文章吗？**
    - ☐ Yes (使用)
    - ☐ No (不使用)
    - ☐ Sometimes (有时使用)

Thank you so much for taking this survey. If you have anything else you want to add you can contact our group via Olivia Steen's WeChatID: "ogsteen" or Lidya Gebremeskel's WeChatID: "lggebremeskel".

非常感谢您今天腾出时间与我们交谈。如果您有任何想要补充的，可以通过 Olivia Steen 的微信：ogsteen 或者 Lidya Gebremeskel 的微信: lggebremeskel 联系我们。

Appendix C: Notification Survey

1. Email:
2. Which of the following would you like to receive notifications from? Check boxes
    - ☐ Google
    - ☐ NCBI
    - ☐ US Patent and Trademark Office
3. If you would like to receive notifications from NCBI, what format would you like the reports in?
    - ☐ Summary
    - ☐ Abstract
4. How often would you like to receive notifications?
    - ☐ Every day
    - ☐ Once a week
    - ☐ Once a month
5. What day of the week would you like to receive these notifications?
    - ☐ Monday
    - ☐ Tuesday
    - ☐ Wednesday
    - ☐ Thursday
    - ☐ Friday
    - ☐ Saturday
    - ☐ Sunday

Appendix D: Specific Keyword Survey

Hello, we are working on a project to find the best research literature options for Hangzhou DAC Biotech, we need some additional data on keywords used during your research. Please provide specific words or phrases you would use to make a search on a literature database when trying to find relevant articles. These terms will be run through different database systems to better understand which database would work best for your research. 谢谢!

1. Name:
2. Email:
3. Department:
    a. Biology
    b. Chemistry
    c. Pharmacology
    d. Other _____
4. Specific Search Words or Terms
   Provide the search terms you use while finding literature
   (example: “Antibody High Mannose” or “Antibody Production”, etc.)

Appendix E: Relevance Survey

Shown in this appendix is a brief example of the type of questions in our relevance survey. At the top of each page of the survey is the search term searched on each of the six databases. Each page consists of six questions containing a picture of the first five results on the database corresponding to the search term. Below the picture is a matrix question on which the respondent is asked to rank the result from the above picture as either relevant, somewhat relevant, or not relevant.

Relevance Survey

Instruction:
This survey will be used to obtain data to gauge the relevance of various literature database options for our research. At the top of each page is the search term that has been run on six different databases. Screen captures of the top five results are shown. For each question, please choose whether or not you think the result is relevant to the search term. In other words, based on what you see in the image, would you click on the result because you think it might contain relevant information? Please answer as many questions as you can, your participation is much appreciated! (Disclaimer: You participation is voluntary, you can stop participating at any time. The results will remain anonymous.)

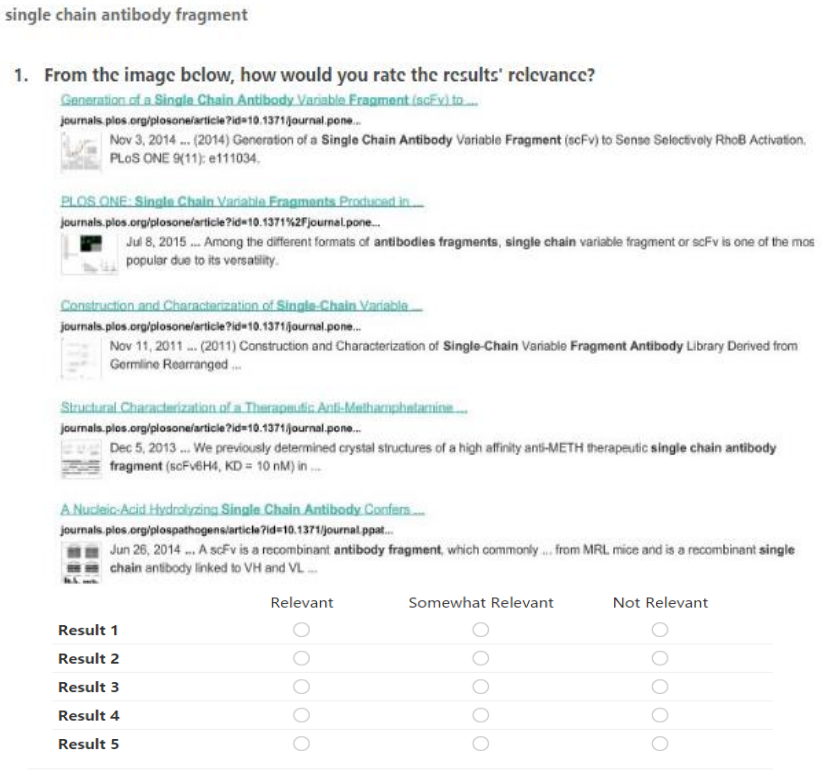


Figure 4. Image of relevance survey

Appendix F: Hangzhou DAC Biotech Employees Interview

Disclaimer: The following responses are summaries, not exact transcripts, of our interviews. These summaries are written by the interviewer and accurately reflect the response and intent of the interviewee.

Interview 1

Q: What types of online resources do you use when searching for research?

A: I prefer to use Google and NCBI.

Q: Why do you choose to use these resources?

A: I like to use Google because it holds lots of different types of information and holds many free articles. However, because Google is blocked by the firewall, the company mostly uses other databases because of the inconvenience.

Q: What are the most convenient?

A: I think NCBI is the most convenient to use. However, most analytical articles need to be purchased from a publisher and cost too much money for us researchers.

Q: Do you use any offline options?

A: We sometimes go to the Zhejiang University Library for hard copy information and sometimes find useful information there.

Q: What do you find difficult about your research process?

A: Because ADC is such a new field, very few people research it. The articles on ADCs are very limited, especially analytical articles.

Q: What suggestions do you have to improve the information retrieval system you use?

A: If the company had a larger budget, we could spend more money on databases. When we do experiments, we compare their results to results of articles we find online.

Q: How do you document research you find?

A: I try to download articles if possible, or take a screenshot or photo and keep it on my personal computer.

Q: How is information shared in the company?

A: Every Monday, there is a company-wide meeting where the researchers present their reports from the past week.

Interview 2

Q: What types of online resources do you use when searching for research?

A: I like to use NCBI, specifically PubMed within NCBI.

Q: Why do you choose to use these resources?

A: NCBI is the only one that has information about the viruses I works with.

Q: Do any resources you use require a subscription?

A: I tend to only searches free literature databases.

Q: What resources do you use to get access to articles behind paywall?

A: If I need something, I ask my friend from university to get articles through the university library system.

Q: Do you use other information sources?

A: I ask other people for help if I am having trouble.

Q: What do you find difficult during the search process?

A: I have difficulty with language barriers when researching literature databases that are in English. However, China has few databases that pertain to biology, and American database websites are the first choice for the company.

Q: Do you have any suggestions to improve data retrieval methods?

A: I think that it could potentially be beneficial for biotech companies to organize a team to buy access to an expensive literature database in order to share the cost.

Interview 3

Q: What types of online resources do you use when searching for research?

A: I like to use Google, but it requires a VPN in China to be used. I also like to use Scifinder and Reaxys. I use these databases because I am a chemist and they contain information more relevant to me for finding information on reactions and chemicals.

Q: Do you have access to any databases through subscriptions?

A: I have my own personal channels for getting information that are not through the company. The company does not have subscriptions that contain the information I am looking for.

Q: Which resource is easiest for you to use?

A: I find that not many databases have been easy to use. Chinese literature databases do not seem to be up to date, and literature written in English can be challenging to read sometimes.

Q: Do you use any other resources to find information?

A: I use the library sometimes, but I would like to have more resources.

Q: What do you find difficult about your research process?

A: I have problems getting full access to literature behind paywalls and most websites only show abstracts of articles.

Interview 4

Q: What types of online resources do you use when searching for research?

A: I like to use NCBI and ScienceDirect the most

Q: Why do you choose to use these resources?

A: NCBI has the most information in the company's field of work. The abstracts are free and some of the full-length papers are total free. I use ScienceDirect because we have a company-wide account for all the researchers to use; however, NCBI is the easiest to find information on.

Q: Do you use any other resources to find information?

A: I seldom use any other resources because most information is online. In China, Google can not be used, so I uses Bing and Yahoo instead even though neither Bing or Yahoo are as useful as Google. Baidu is very good at finding results in Chinese, but not in English, and is also far worse that google in my opinion.

Q: What do you find difficult about your research process?

A: It is easy to find information you are looking for, but usually the articles are behind paywalls and I can only read the abstract.

Interview 5

Q: What types of online resources do you use when searching for research?

A: I primarily use Google, NCBI, and the FDA as resources.

Q: Why do you choose to use these resources?

A: I like to use these resources because they are high-quality, and it is easy to find the information I am looking for while using them. I like to use PubMed and Google the most, even though Google requires a VPN.

Q: How often do you look for research?

A: For my field, data does not come out very fast, so I do not have many problems getting up-to-date information. I normally look for information twice a week.

Q: What do you find difficult about your research process?

A: Because some websites are blocked, I sometimes have to use VPN, which can be unstable and make it hard to do research. I also have trouble obtaining the full texts for articles that are behind paywalls because the company subscription to ScienceDirect does not have full coverage of the information I need.

Appendix G: Hangzhou Dianzi University Librarian Interview Summary

Disclaimer: The following responses are summaries, not exact transcripts, of our interviews. These summaries are written by the interviewer and accurately reflect the response and intent of the interviewee.

HDU subscribes to many research databases, such as Web of Science and ScienceDirect. They have a VPN that one can use to access these resources off campus, however it is only available to teachers, so non-HDU personnel can not use it. They may have tried letting other entities outside of HDU use the VPN, but are no longer allowing it due to potentially violating terms of a license agreement.

Appendix H: Worcester Polytechnic Institute Librarian Summary

Disclaimer: The following responses are summaries, not exact transcripts, of our interviews. These summaries are written by the interviewer and accurately reflect the response and intent of the interviewee.

WPI subscribes to a wide variety of literature databases. They are not legally allowed to share or sublease any of their subscriptions to any other organizations because they get their subscriptions through various school consortia they are a part of as well as the database licensing agreement. WPI spends approximately 400,000 USD per year on their ScienceDirect subscription which allows access to all journals in the database. They recommended looking into the Directory of Open Access Journals as an open access research literature resource for Hangzhou DAC Biotech.

Appendix I: Sponsor Interview

Disclaimer: The following responses are summaries, not exact transcripts, of our interviews. These summaries are written by the interviewer and accurately reflect the response and intent of the interviewee.

Q: What kind of relationships with other universities and organizations do you have?

A: The company collaborates with a Chinese university and 2 other companies. The connections between us are not very strong and reasonably new, and it would not be possible to share resources with them.

Q: Does your company subscribe to any literature databases?

A: The company has a 400 USD yearly subscription to ScienceDirect that includes access to a handful of relevant journals on biotechnology. There is a government website that provides some access to articles behind a paywall, but it only includes about 50% of what we need, the problem is getting access to the other 50%. Even with this subscription, the company still pays per article for most Elsevier publications. Usually the price of articles ranges between 30 and 40 USD.

Q: What is your budget for research resources?

A: The company budget for research is very limited. Because the company is so small, we can not afford to spend too much on expensive subscriptions. We are mostly focused on moving our ADCs into clinical trials within the next three years. Because of this, it is not uncommon for researchers here to fall behind on research because they forget to keep searching databases. A notification system would be a huge help.

Appendix J: Employee Survey Results

Table 6. Employee survey results

	Survey 1	Survey 2	Survey 3	Survey 4	Survey 5
Frequency of conducting research	Twice a Week	Once a Week	Twice a Week	Once a Week	Less than Twice a Week
Databases used when researching	ScienceDirect, NCBI, Google, Google Scholar, CFDA	ScienceDirect, Google, SciFinder, Reaxys, Web of Science/Web of Knowledge	SciFinder, CFDA, Web of Science/ Web of Knowledge, FDA	ScienceDirect, Google, SciFinder, Reaxys, Web of Science/Web of Knowledge	NCBI, CFDA, FDA
Other databases used	Baidu Xeushu, CNKI, Wangfang Data, Springer Linker, C-Quip				
Favorite platform	NCBI	SciFinder	CFDA, FDA, Web of Science	SciFinder	NCBI
Trouble accessing full-length papers	Yes	Yes	Yes	Yes	Yes
Trouble searching for research	No	No	No	No	Yes
Frequency of trouble finding research					Sometimes
Trouble keeping up-to-date	Yes	Yes	Yes	Yes	Yes
Use notification system	No	No	No	No	No
Frequency of spending money on papers not covered by subscription	Rarely	Rarely	Rarely	Rarely	Rarely

	Survey 6	Survey 7	Survey 8	Survey 9	Survey 10
Frequency of conducting research	Twice a Week	Once a Week	Once a Week	Less than Twice a Week	Once a Week
Databases used when researching	Science Direct, NCBI, Google, Google Scholar, CFDA, FDA	Science Direct, NCBI, Google, Google Scholar, ResearchGate	Science Direct, NCBI, Google, Google Scholar, FDA	NCBI, Google, Google Scholar, CFDA, FDA	SciFinder
Other databases used	Baidu Xeushu, CNKI, Wangfang Data, Springer Linker, C-Quip		Baidu Xeushu		
Favorite platform	NCBI	NCBI	NCBI	NCBI, FDA, CFDA	SciFinder
Trouble accessing full-length papers	Yes	Yes	Yes	Yes	Yes
Trouble searching for research	No	Yes	Yes	No	Yes
Frequency of trouble finding research		Most of the time	Sometimes		Rarely
Trouble keeping up-to-date	Yes	No	Yes	No	Yes
Use notification system	No	No	No	No	No
Frequency of spending money on papers not covered by subscription	Rarely	Rarely	Rarely	Never	Most of the time

	Survey 11	Survey 12	Survey 13	Survey 14	Survey 15
Frequency of conducting research	Every Day	Once a Week		Less than Twice a Week	Twice a Week
Databases used when researching	NCBI, Google, Google Scholar	NCBI, ResearchGate, FDA	Google, SciFinder, Reaxys, Web of Science/Web of Knowledge	Science Direct, NCBI, Google, CFDA, FDA	Science Direct, NCBI, Google Scholar, SciFinder, ResearchGate, Academia, FDA
Other databases used		UCSC, Baidu Xeushu, CNKI	ACS, RSC, ElSevier		Springer Linker
Favorite platform	NCBI	NCBI	Web of Science, Google	NCBI	Other
Trouble accessing full-length papers	Yes	Yes	Yes	Yes	Yes
Trouble searching for research	Yes	Yes	Yes	Yes	Yes
Frequency of trouble finding research	Sometimes	Most of the time	Sometimes	Rarely	Most of the time
Trouble keeping up-to-date	No	No	No	Yes	Yes
Use notification system	No	No	No	No	No
Frequency of spending money on papers not covered by subscription	Rarely	Always	Sometimes	Sometimes	Never

	Survey 16	Survey 17	Survey 18	Survey 19	Survey 20
Frequency of conducting research	Twice a Week	Once a Week	Less than Twice a Week	Every Day	Once a Week
Databases used when researching	NCBI, Google, Google Scholar	SciFinder	Google, Web of Science/Web of Knowledge	NCBI	NCBI, Google, CFDA
Other databases used					
Favorite platform	NCBI	SciFinder		CNKI, NCBI	Google
Trouble accessing full-length papers	Yes	Yes	Yes	No	Yes
Trouble searching for research	No	Yes	Yes	Yes	Yes
Frequency of trouble finding research		Sometimes	Most of the time	Rarely	Sometimes
Trouble keeping up-to-date	No	No	No	No	No
Use notification system	No	No	No	No	No
Frequency of spending money on papers not covered by subscription	Rarely	Never	Always	Rarely	Most of the time

	Survey 21	Survey 22
Frequency of conducting research	Once a Week	Once a Week
Databases used when researching	NCBI, Google Scholar, CFDA	Google, Google Scholar, CFDA
Other databases used		
Favorite platform	NCBI, Google	Google
Trouble accessing full-length papers	Yes	No
Trouble searching for research	Yes	No
Frequency of trouble finding research	Sometimes	
Trouble keeping up-to-date	Yes	No
Use notification system	No	No
Frequency of spending money on papers not covered by subscription	Never	Rarely

Appendix K: Notification Survey Results

Table 7. Notification form results

	System	NCBI format (if applicable)	Frequency of alert	Day of alert (if applicable)
Form 1	NCBI	Abstract	Every day	
Form 2	NCBI	Summary	Once a week	Monday
Form 3	NCBI	Summary	Once a week	Tuesday
Form 4	NCBI	Abstract	Once a week	Monday
Form 5	NCBI	Abstract	Once a week	Monday
Form 6	NCBI	Abstract	Once a week	Monday
Form 7	NCBI	Abstract	Once a month	Tuesday
Form 8	NCBI	Abstract	Once a week	Wednesday
Form 9	NCBI	Summary	Once a week	Monday
Form 10	Google, NCBI	Once a week		Monday
Form 11	Google, NCBI, USPTO	Abstract	Every day	Friday
Form 12	Google, NCBI, USPTO	Summary, Abstract	Every day	Monday, Friday, Saturday, Sunday
Form 13	Google, NCBI, USPTO	Abstract	Once a week	Monday
Form 14	USPTO	Abstract	Once a week	Friday
Form 15	NCBI, USPTO	Abstract	Once a week	Monday
Form 16	NCBI, USPTO	Summary	Once a week	Tuesday
Form 17	Google	Abstract	Every day	Monday
Form 18	Google		Once a week	
Form 19	Google, NCBI	Summary	Once a week	Monday

Appendix L: Specific Keyword Survey Results

Table 8. Advanced keyword survey results

Biology	Chemistry	Pharmacology
Antibody drug conjugate quality assurance	Organic synthesis hydrolysis	Antibody drug conjugate AND pharmacology
Antibody drug conjugate formulation	Biological mechanisms of chemicals	Antibody drug conjugate AND pharmacokinetics
Antibody drug conjugate AND prostate cancer	Drug quality control	Antibody drug conjugate AND clinical trial
Antibody drug conjugate AND multiple myeloma	Recent developments of antibody drug conjugate	Preclinical antibody drug conjugate
Stable antibody expression	Polymerisation separation	Safety evaluation antibody drug conjugate
Transient antibody expression		Antibody drug conjugate quality assurance
Stable expression vector		Pharmacokinetic assays of antibody drug conjugate
Antibody high mannose		
Antibody acidic charge variant		
Antibody basic charge variant		
Bispecific antibody		
Single chain antibody fragment		
Antibody panning		
Antibody humanization		
Antibody expression		
CHO cell screening		
Stable cell establishment		
Cell culture		
Medium antibody expression technology development		
Stable transfection		
Glycosylation of antibody		
Monoclonal antibody expression		
Pharmacokinetic assays of antibody drug conjugate		

Appendix M: Relevance Survey Results

Table 9. Relevance survey results

Biology Keywords (Survey 1-3):

Survey 1

PLOS Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC quality assurance	0	0	0	0.5	0.5
ADC formulation	0.5	0.5	0.5	0.5	0.5
ADC AND prostate cancer	0.5	0.5	0.5	0.5	0.5
ADC AND multiple myeloma	0.5	0.5	0.5	0.5	0
single chain antibody fragment	1	1	1	1	1
glycosylation of antibody	1	1	1	1	1
monoclonal antibody expression	0	0	0	0	0
pharmacokinetic assays of ADC	0.5	0.5	1	1	0.5

Survey 1

DOAJ Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC quality assurance	0	0	0	0	0
ADC formulation	0	0	0	0	0
ADC AND prostate cancer	0	0	0	0	0
ADC AND multiple myeloma	0	0	0	0	0
single chain antibody fragment	1	1	1	1	1
glycosylation of antibody	0	0	0	0.5	0
monoclonal antibody expression	0	0	0	0.5	0
pharmacokinetic assays of ADC	0	0	0	0	0

Survey 1

DeepDyve Results	**Result 1**	**Result 2**	**Result 3**	**Result 4**	**Result 5**
ADC quality assurance	0	0	0	0	0
ADC formulation	0.5	0.5	1	0.5	0.5
ADC AND prostate cancer	1	1	1	1	1
ADC AND multiple myeloma	0.5	0.5	0.5	0.5	0.5
single chain antibody fragment	1	1	1	1	1
glycosylation of antibody	0.5	0.5	0.5	0.5	1
monoclonal antibody expression	0	0	0	0	1
pharmacokinetic assays of ADC	0.5	0.5	1	0.5	0.5

Survey 1

ScienceDirect Results	**Result 1**	**Result 2**	**Result 3**	**Result 4**	**Result 5**
ADC quality assurance	0.5	0.5	0.5	0.5	0
ADC formulation	0.5	0.5	0.5	0.5	0.5
ADC AND prostate cancer	0.5	0.5	0.5	0.5	0.5
ADC AND multiple myeloma	0	0.5	1	0.5	0.5
single chain antibody fragment	1	1	1	1	1
glycosylation of antibody	0.5	1	1	0.5	0.5
monoclonal antibody expression	0	0	0.5	0	0.5
pharmacokinetic assays of ADC	1	0.5	0.5	0.5	0

Survey 1

NCBI Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC quality assurance	0.5	1	0.5	0.5	1
ADC formulation	0.5	0.5	0.5	0.5	0.5
ADC AND prostate cancer	0.5	0.5	0.5	0.5	0.5
ADC AND multiple myeloma	0.5	0.5	0	0	0
single chain antibody fragment	1	1	1	1	0.5
glycosylation of antibody	0	1	0.5	0	0.5
monoclonal antibody expression	0	0	0	0	0
pharmacokinetic assays of ADC	0.5	0.5	0.5	0.5	0.5

Survey 1

Google Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC quality assurance	0.5	1	0.5	0.5	1
ADC formulation	0.5	1	0	0.5	1
ADC AND prostate cancer	0.5	0.5	0.5	0.5	0.5
ADC AND multiple myeloma	0.5	0.5	0.5	0.5	0
single chain antibody fragment	1	1	1	1	1
glycosylation of antibody	1	0.5	1	0	0.5
monoclonal antibody expression	0.5	0.5	0	0	1
pharmacokinetic assays of ADC	1	1	1	0	1

Survey 2

PLOS Results	Result 1	Result 2	Result 3	Result 4	Result 5
stable antibody expression	0.5	0	0	0.5	1
stable expression vector	0	0.5	0	0.5	1
antibody high mannose	0.5	0.5	0	0.5	0
bispecific antibody	0.5	1	0.5	0.5	0.5
single chain antibody fragment	1	1	1	1	1
antibody panning	0.5	0.5	0.5	1	0.5
antibody humanization	0	1	1	0.5	0.5
antibody expression	0.5	0.5	0.5	1	0.5
stable cell establishment	0	0	0.5	0	0.5
stable transfection	0.5	0	0	0	0

Survey 2

DOAJ Results	Result 1	Result 2	Result 3	Result 4	Result 5
stable antibody expression	0	0	0.5	0	0.5
stable expression vector	0	0.5	0.5	0	0.5
antibody high mannose	1	0.5	0.5	0.5	0
bispecific antibody	0.5	0.5	0.5	0	0.5
single chain antibody fragment	0	0	0	0.5	0.5
antibody panning	0	0.5	0	0	0
antibody humanization	0.5	0.5	1	0.5	1
antibody expression	0	0	0.5	0.5	0
stable cell establishment	0	0	0.5	0	0.5
stable transfection	0	0.5	0.5	0	0

Survey 2

DeepDyve Results	**Result 1**	**Result 2**	**Result 3**	**Result 4**	**Result 5**
stable antibody expression	1	0.5	0.5	0.5	1
stable expression vector	0.5	0.5	0	0	0.5
antibody high mannose	0	0.5	0.5	1	0.5
bispecific antibody	1	0.5	0.5	1	0.5
single chain antibody fragment	0.5	1	1	0.5	0
antibody panning	0.5	0.5	0	0	0.5
antibody humanization	0.5	1	1	1	0.5
antibody expression	1	1	1	0.5	0
stable cell establishment	0	0	0	0.5	0
stable transfection	0	0	0.5	0	0

Survey 2

ScienceDirect Results	**Result 1**	**Result 2**	**Result 3**	**Result 4**	**Result 5**
stable antibody expression	1	0.5	1	0.5	0.5
stable expression vector	0.5	0.5	0.5	0.5	1
antibody high mannose	1	0.5	0.5	0.5	0.5
bispecific antibody	0.5	1	0.5	0.5	0.5
single chain antibody fragment	0.5	0.5	0.5	1	1
antibody panning	0.5	0.5	1	0.5	0.5
antibody humanization	0	0	0.5	1	0.5
antibody expression	0.5	0.5	0.5	1	0.5
stable cell establishment	0	0	0.5	0	0.5
stable transfection	0.5	1	0.5	0	0.5

Survey 2

NCBI Results	Result 1	Result 2	Result 3	Result 4	Result 5
stable antibody expression	0	0.5	0	0.5	0.5
stable expression vector	0.5	0.5	0	0	0.5
antibody high mannose	0	0.5	0	0	0.5
bispecific antibody	0.5	0.5	0	0	0.5
single chain antibody fragment	0.5	1	1	0.5	0
antibody panning	0.5	0.5	1	1	0.5
antibody humanization	0.5	1	0.5	0.5	1
antibody expression	0.5	0.5	0	0	0
stable cell establishment	0	0.5	0	0.5	0
stable transfection	0	0.5	0	0	0.5

Survey 2

Google Results	Result 1	Result 2	Result 3	Result 4	Result 5
stable antibody expression	1	0.5	0.5	0.5	0
stable expression vector	0.5	0.5	1	0	0.5
antibody high mannose	0	0.5	1	0.5	1
bispecific antibody	0.5	0.5	0.5	1	1
single chain antibody fragment	0.5	0.5	1	0.5	0.5
antibody panning	0	0.5	1	0.5	1
antibody humanization	0.5	0.5	1	1	0.5
antibody expression	0.5	0.5	1	0.5	0.5
stable cell establishment	0	0	0	0	0.5
stable transfection	0	0.5	0.5	0	0.5

Survey 3

PLOS Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC quality assurance	0.5	1	0	1	1
ADC formulation	0.5	0.5	0.5	0.5	0.5
ADC AND prostate cancer	1	0	0	0.5	0.5
ADC AND multiple myeloma	0.5	1	1	0.5	0.5
glycosylation of antibody	1	1	1	1	1
monoclonal antibody expression	0	1	0.5	0.5	0.5
pharmacokinetic assays of ADC	1	1	0.5	1	1

Survey 3

DOAJ Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC quality assurance	0	0	0	0	0
ADC formulation	0	0	0	0	0
ADC AND prostate cancer	1	0	0	0	0
ADC AND multiple myeloma	0.5	0	0	0	0
glycosylation of antibody	0	1	0.5	0.5	0.5
monoclonal antibody expression	0	0	0.5	0	0
pharmacokinetic assays of ADC	0	0	0	0	0

Survey 3

DeepDyve Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC quality assurance	0	0	0	0	0
ADC formulation	1	1	1	0.5	1
ADC AND prostate cancer	1	0.5	1	0.5	0.5
ADC AND multiple myeloma	0	1	1	1	1
glycosylation of antibody	0.5	0.5	0.5	0.5	1
monoclonal antibody expression	1	0.5	0.5	1	1
pharmacokinetic assays of ADC	1	1	1	0.5	0.5

Survey 3

ScienceDirect Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC quality assurance	1	0.5	0	0	0
ADC formulation	1	0.5	1	1	1
ADC AND prostate cancer	1	0	0.5	1	1
ADC AND multiple myeloma	0	0	0.5	0.5	0
glycosylation of antibody	1	0.5	1	0.5	0
monoclonal antibody expression	0.5	0	0	0	0.5
pharmacokinetic assays of ADC	0	0.5	0.5	0.5	0.5

Survey 3

NCBI Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC quality assurance	1	0	1	1	0.5
ADC formulation	1	0.5	0.5	1	0
ADC AND prostate cancer	1	1	0	1	1
ADC AND multiple myeloma	1	1	1	0.5	0.5
glycosylation of antibody	0	1	0.5	0	0.5
monoclonal antibody expression	0	0	0	0	0
pharmacokinetic assays of ADC	1	0.5	0.5	0	0.5

Survey 3

Google Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC quality assurance	0	0.5	1	0.5	1
ADC formulation	1	1	1	0.5	1
ADC AND prostate cancer	0.5	0.5	0	0	1
ADC AND multiple myeloma	0.5	1	0.5	1	0
glycosylation of antibody	1	0.5	1	0	0.5
monoclonal antibody expression	0.5	0.5	0.5	0.5	1
pharmacokinetic assays of ADC	1	1	1	1	1

Biology Keywords (Survey 4-5):

Survey 4

PLOS Results	**Result 1**	**Result 2**	**Result 3**	**Result 4**	**Result 5**
organic synthesis hydrolysis	0.5	0	0.5	0	0
biological mechanisms of chemicals	0	0.5	0.5	0	1
drug quality control	0.5	0	0	1	0
recent developments of antibody drug conjugate	0.5	0	0	1	0

Survey 4

DOAJ Results	**Result 1**	**Result 2**	**Result 3**	**Result 4**	**Result 5**
organic synthesis hydrolysis	0	1	0.5	1	0
biological mechanisms of chemicals	0.5	0	0	1	0.5
drug quality control	1	0.5	0	0.5	0.5
recent developments of antibody drug conjugate	0	0	0	0	0

Survey 4

DeepDyve Results	**Result 1**	**Result 2**	**Result 3**	**Result 4**	**Result 5**
organic synthesis hydrolysis	0	0.5	1	0	0.5
biological mechanisms of chemicals	0	0	1	0.5	0
drug quality control	0.5	0.5	0	0	0.5
recent developments of antibody drug conjugate	1	1	0.5	1	0.5

Survey 4

ScienceDirect Results	Result 1	Result 2	Result 3	Result 4	Result 5
organic synthesis hydrolysis	0	0.5	0	0	0
biological mechanisms of chemicals	0	0.5	0	1	0
drug quality control	0.5	0	0	0	0.5
recent developments of antibody drug conjugate	1	1	1	1	1

Survey 4

NCBI Results	Result 1	Result 2	Result 3	Result 4	Result 5
organic synthesis hydrolysis	0.5	0	0.5	0	0
biological mechanisms of chemicals	0	0	0.5	0	0.5
drug quality control	0.5	0	0	1	0.5
recent developments of antibody drug conjugate	1	0.5	0.5	1	1

Survey 4

Google Results	Result 1	Result 2	Result 3	Result 4	Result 5
organic synthesis hydrolysis	0	0.5	0	1	1
biological mechanisms of chemicals	0	0.5	0	1	0.5
drug quality control	0.5	0.5	0	0	0.5
recent developments of antibody drug conjugate	0.5	1	1	0.5	1

Survey 5

PLOS Results	Result 1	Result 2	Result 3	Result 4	Result 5
organic synthesis hydrolysis	1	1	0.5	0	1
biological mechanisms of chemicals	1	0.5	0.5	0	0
recent developments of antibody drug conjugate	0.5	0	0	0	0

Survey 5

DOAJ Results	Result 1	Result 2	Result 3	Result 4	Result 5
organic synthesis hydrolysis	0.5	0.5	0	0.5	1
biological mechanisms of chemicals	0	0.5	0	0	0
recent developments of antibody drug conjugate	0	0	0	0	0

Survey 5

DeepDyve Results	Result 1	Result 2	Result 3	Result 4	Result 5
organic synthesis hydrolysis	1	1	0	0	0
biological mechanisms of chemicals	0.5	0.5	0.5	0	0
recent developments of antibody drug conjugate	1	1	1	1	0.5

Survey 5

ScienceDirect Results	Result 1	Result 2	Result 3	Result 4	Result 5
organic synthesis hydrolysis	0	0	0	0.5	0
biological mechanisms of chemicals	0	0	0	0.5	0
recent developments of antibody drug conjugate	1	1	1	1	1

Survey 5

NCBI Results	Result 1	Result 2	Result 3	Result 4	Result 5
organic synthesis hydrolysis	0.5	0	0.5	0	0
biological mechanisms of chemicals	0	0	0	0	0
recent developments of antibody drug conjugate	1	0.5	0.5	0.5	0.5

Survey 5

Google Results	**Result 1**	**Result 2**	**Result 3**	**Result 4**	**Result 5**
organic synthesis hydrolysis	0.5	0	0.5	1	0.5
biological mechanisms of chemicals	1	1	0.5	0.5	1
recent developments of antibody drug conjugate	1	1	1	1	0.5

Pharmacology Keywords (Survey 6-7):

Survey 6

PLOS Results	**Result 1**	**Result 2**	**Result 3**	**Result 4**	**Result 5**
ADC AND pharmacology	1	1	1	1	1
ADC AND clinical trial	1	1	1	0.5	0.5
preclinical Antibody Drug Conjugate	1	1	1	1	1
safety evaluation Antibody Drug Conjugate	1	1	1	1	1
Antibody Drug Conjugate quality assurance	1	1	1	1	1

Survey 6

DOAJ Results	**Result 1**	**Result 2**	**Result 3**	**Result 4**	**Result 5**
ADC AND pharmacology	1	1	1	1	1
ADC AND clinical trial	1	1	1	1	1
preclinical Antibody Drug Conjugate	1	1	0.5	1	0
safety evaluation Antibody Drug Conjugate	1	0	0	0	0
Antibody Drug Conjugate quality assurance	0	0	0	0	0

Survey 6

DeepDyve Results	**Result 1**	**Result 2**	**Result 3**	**Result 4**	**Result 5**
ADC AND pharmacology	1	1	1	1	1
ADC AND clinical trial	1	1	1	1	1
preclinical Antibody Drug Conjugate	1	1	1	1	1
safety evaluation Antibody Drug Conjugate	1	1	1	1	1
Antibody Drug Conjugate quality assurance	0	1	1	1	1

Survey 6

ScienceDirect Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC AND pharmacology	1	1	1	1	1
ADC AND clinical trial	1	1	1	1	1
preclinical Antibody Drug Conjugate	1	1	1	1	1
safety evaluation Antibody Drug Conjugate	1	1	1	1	1
Antibody Drug Conjugate quality assurance	0	1	1	1	1

Survey 6

NCBI Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC AND pharmacology	1	1	1	1	1
ADC AND clinical trial	1	1	0.5	1	0.5
preclinical Antibody Drug Conjugate	1	1	1	1	1
safety evaluation Antibody Drug Conjugate	1	1	1	1	1
Antibody Drug Conjugate quality assurance	1	0.5	1	1	1

Survey 6

Google Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC AND pharmacology	1	1	1	1	1
ADC AND clinical trial	1	1	1	1	1
preclinical Antibody Drug Conjugate	1	1	1	1	1
safety evaluation Antibody Drug Conjugate	1	1	1	1	1
Antibody Drug Conjugate quality assurance	1	1	1	0.5	0.5

Survey 7

PLOS Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC AND pharmacology	1	1	1	0	0
Antibody Drug Conjugate AND pharmacokinetics	0.5	0.5	0.5	1	0.5
ADC AND clinical trial	0.5	0.5	0.5	0.5	0
preclinical Antibody Drug Conjugate	1	0.5	1	1	0
safety evaluation Antibody Drug Conjugate	1	1	1	0.5	1
Antibody Drug Conjugate quality assurance	1	1	0	0.5	0.5

Survey 7

DOAJ Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC AND pharmacology	0.5	0.5	1	0	0
Antibody Drug Conjugate AND pharmacokinetics	0	0	0	0	0
ADC AND clinical trial	1	0.5	0.5	0.5	0.5
preclinical Antibody Drug Conjugate	1	0	0	0	0
safety evaluation Antibody Drug Conjugate	1	1	1	0.5	1
Antibody Drug Conjugate quality assurance	0	0	0	0	0

Survey 7

DeepDyve Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC AND pharmacology	1	1	1	1	1
Antibody Drug Conjugate AND pharmacokinetics	1	1	1	1	1
ADC AND clinical trial	1	0.5	0.5	1	0.5
preclinical Antibody Drug Conjugate	1	1	1	1	1
safety evaluation Antibody Drug Conjugate	1	1	1	1	1
Antibody Drug Conjugate quality assurance	0.5	0	0	0	0

Survey 7

ScienceDirect Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC AND pharmacology	1	1	1	1	1
Antibody Drug Conjugate AND pharmacokinetics	1	1	0.5	0	1
ADC AND clinical trial	1	0.5	0.5	1	0.5
preclinical Antibody Drug Conjugate	1	1	1	1	1
safety evaluation Antibody Drug Conjugate	1	1	1	1	0.5
Antibody Drug Conjugate quality assurance	0.5	0	0	0	0

Survey 7

NCBI Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC AND pharmacology	1	1	1	0	0
Antibody Drug Conjugate AND pharmacokinetics	0	0.5	0.5	1	1
ADC AND clinical trial	0.5	0.5	1	0.5	0.5
preclinical Antibody Drug Conjugate	1	1	1	1	1
safety evaluation Antibody Drug Conjugate	0.5	1	1	1	1
Antibody Drug Conjugate quality assurance	0.5	0	0.5	0	0

Survey 7

Google Results	Result 1	Result 2	Result 3	Result 4	Result 5
ADC AND pharmacology	1	1	1	0.5	0.5
Antibody Drug Conjugate AND pharmacokinetics	0	0	0.5	0.5	0
ADC AND clinical trial	1	1	0.5	0.5	0.5
preclinical Antibody Drug Conjugate	1	1	0.5	1	1
safety evaluation Antibody Drug Conjugate	0.5	0.5	1	0.5	0
Antibody Drug Conjugate quality assurance	1	1	0.5	0.5	0.5

Appendix N: Notification Guide

(Begins on next page)

NCBI Notification Tutorial

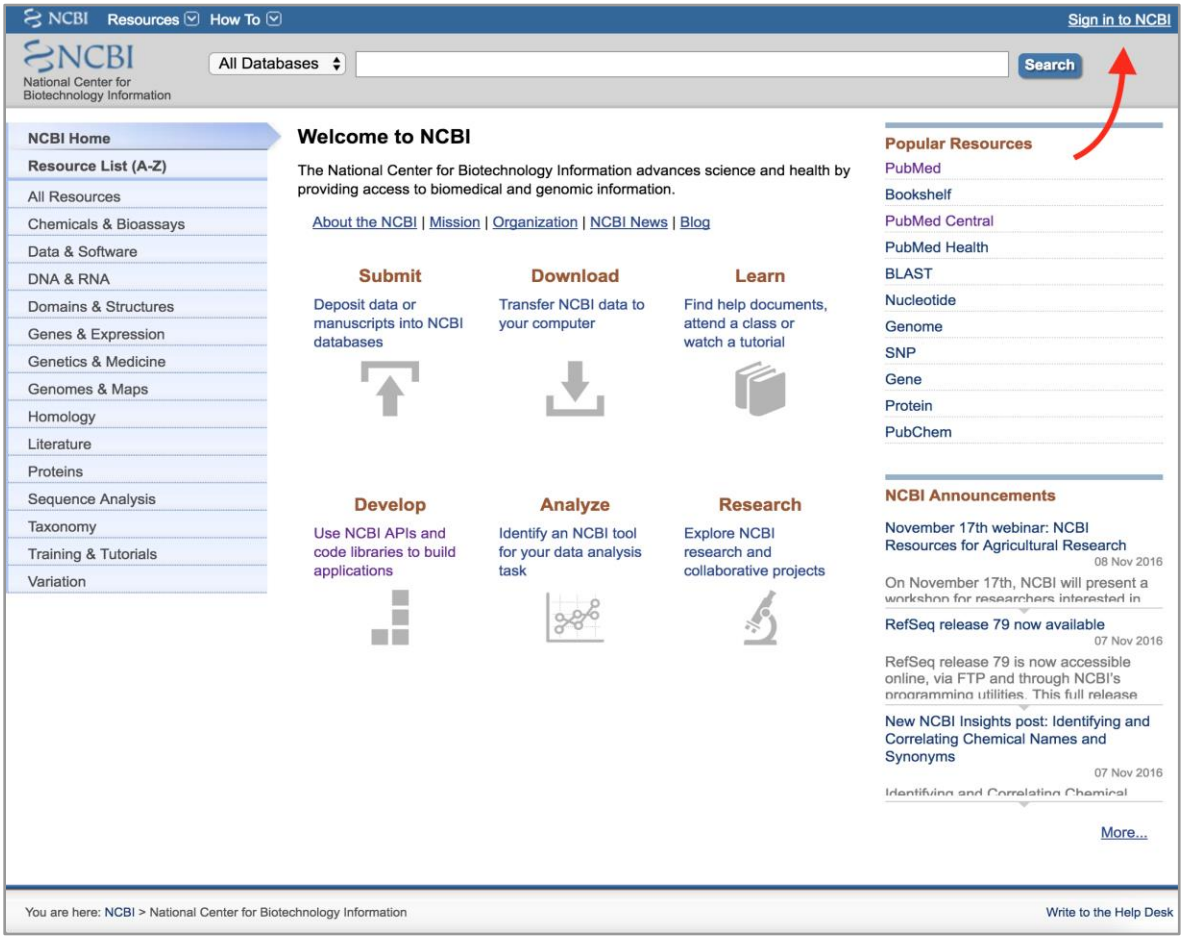
Guide to Setting Up NCBI Alerts

Prepared by Database Team

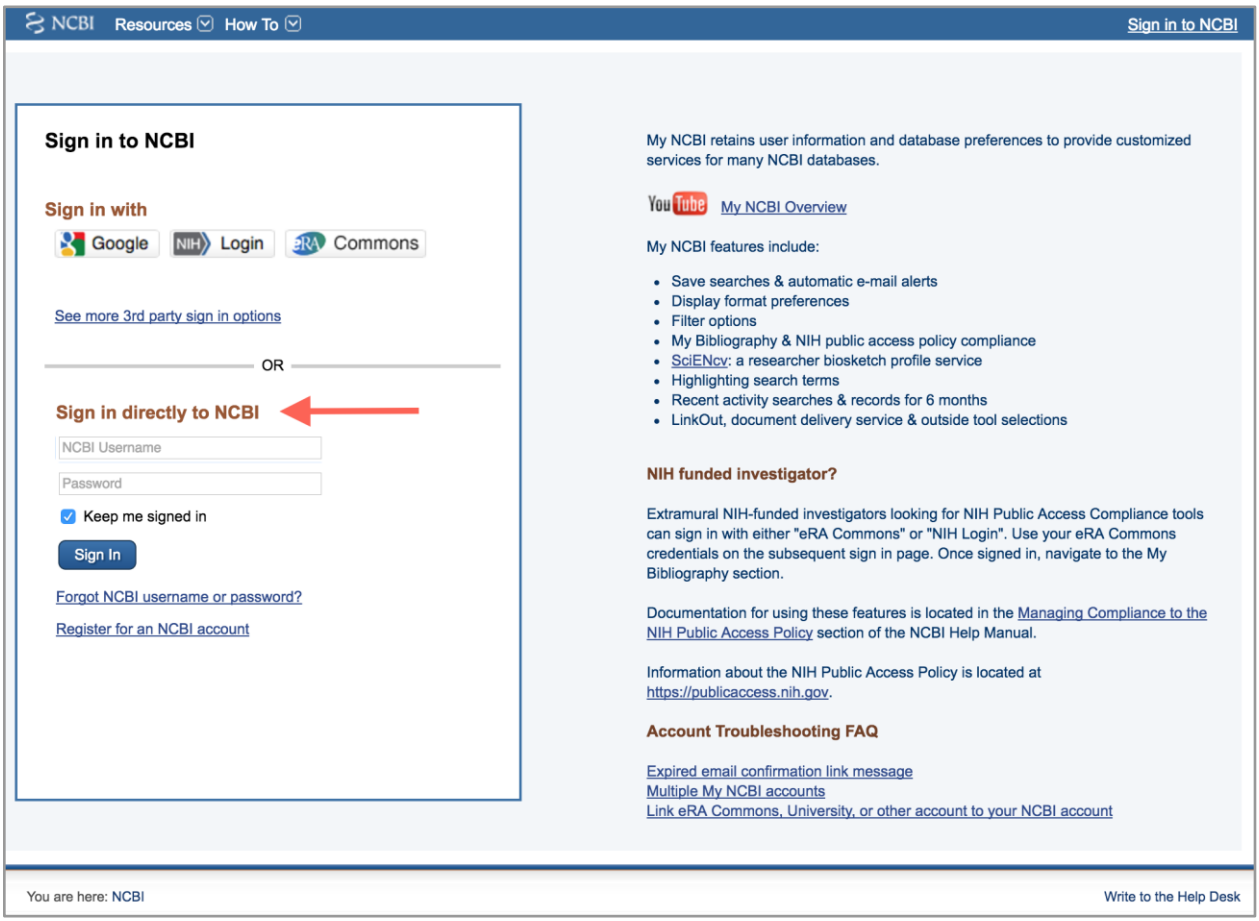
Step 1: Navigate to NCBI's website - https://www.ncbi.nlm.nih.gov/

Step 2: Log into your account

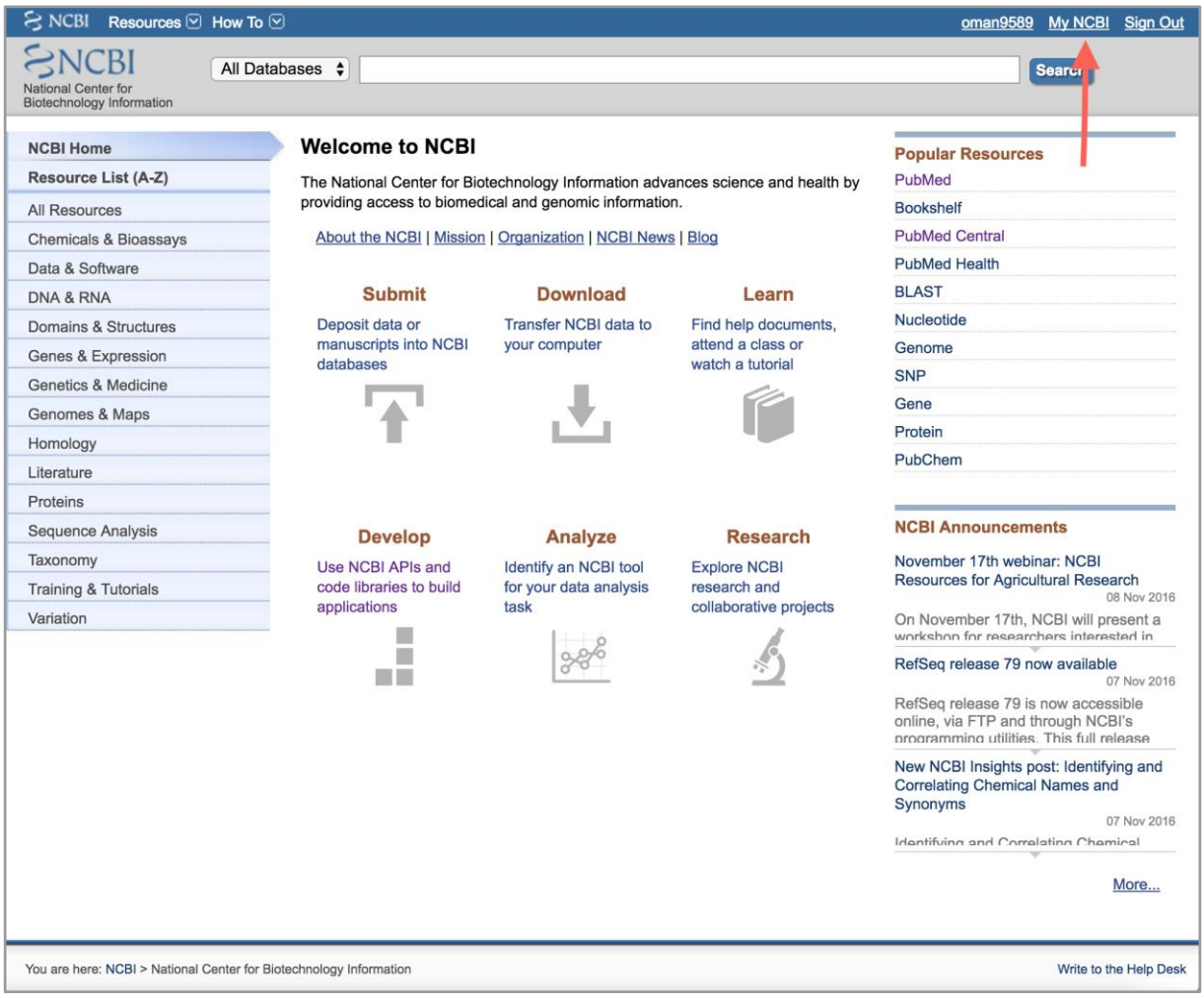
1. Sign in to NCBI (Skip to part 3 of this step if already logged in)



2. Enter your username and password to sign in



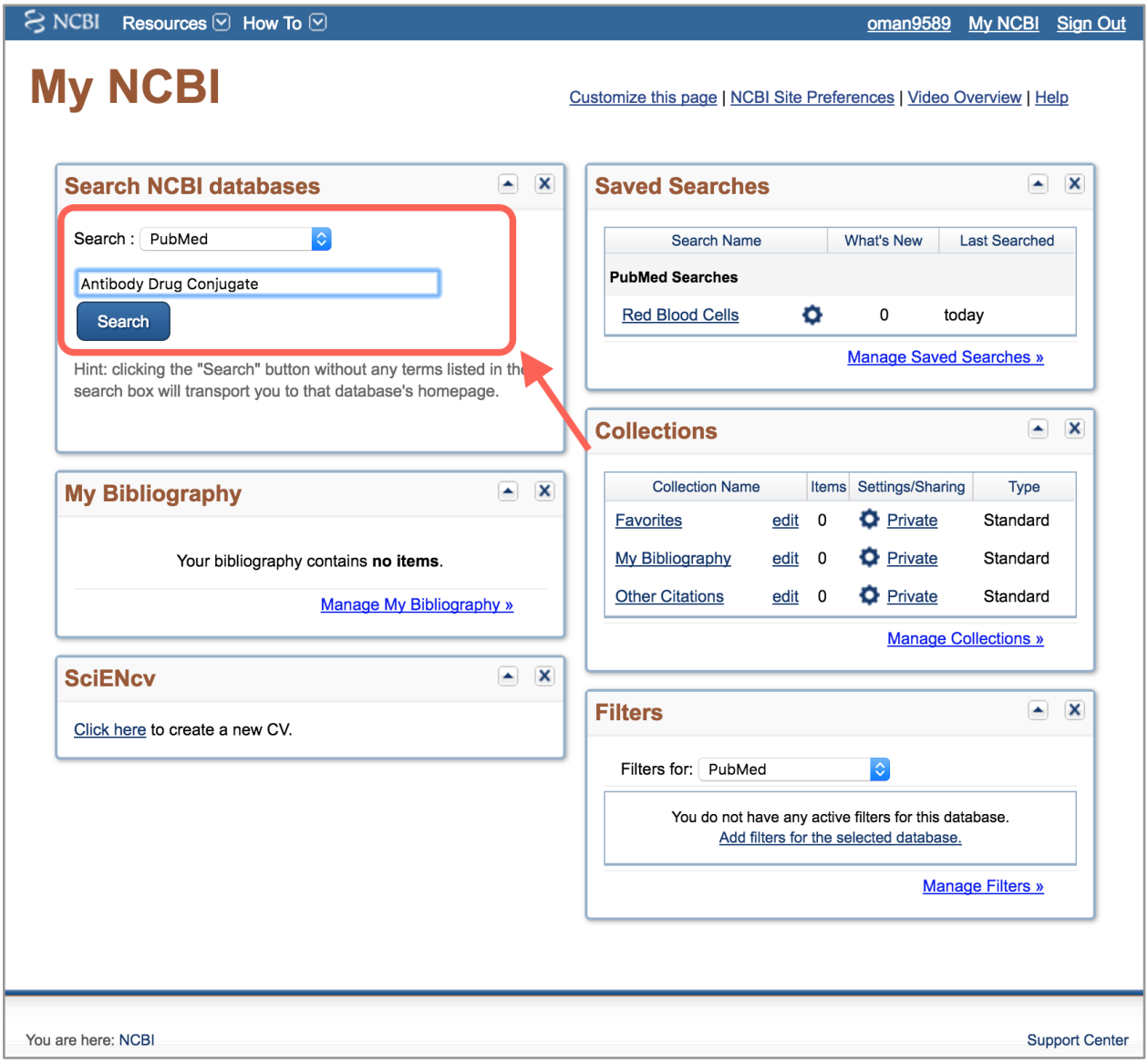
3. Proceed to ‘My NCBI’



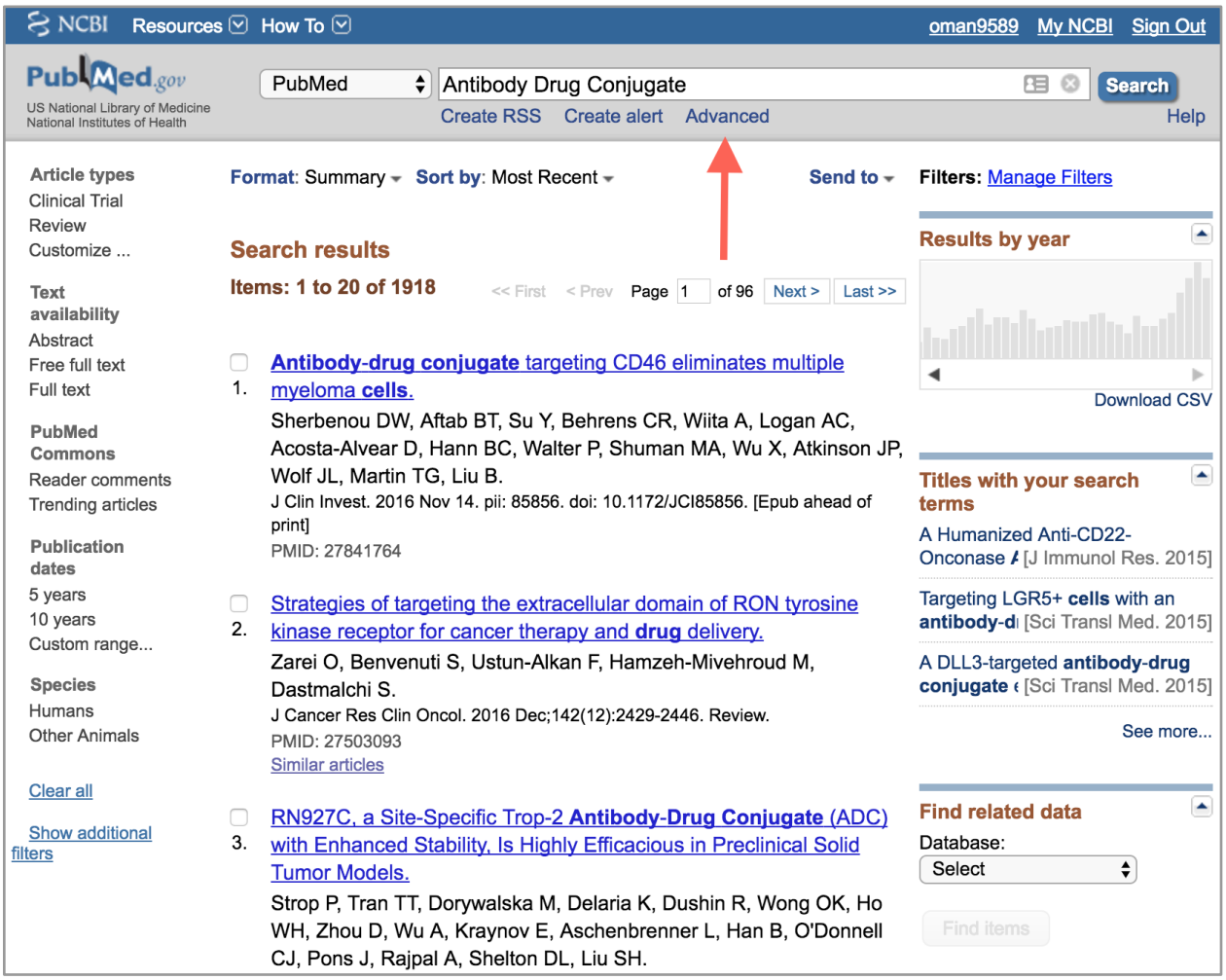
Step 3: Building and adding a saved search

Skip this step to learn about managing and removing saved searches

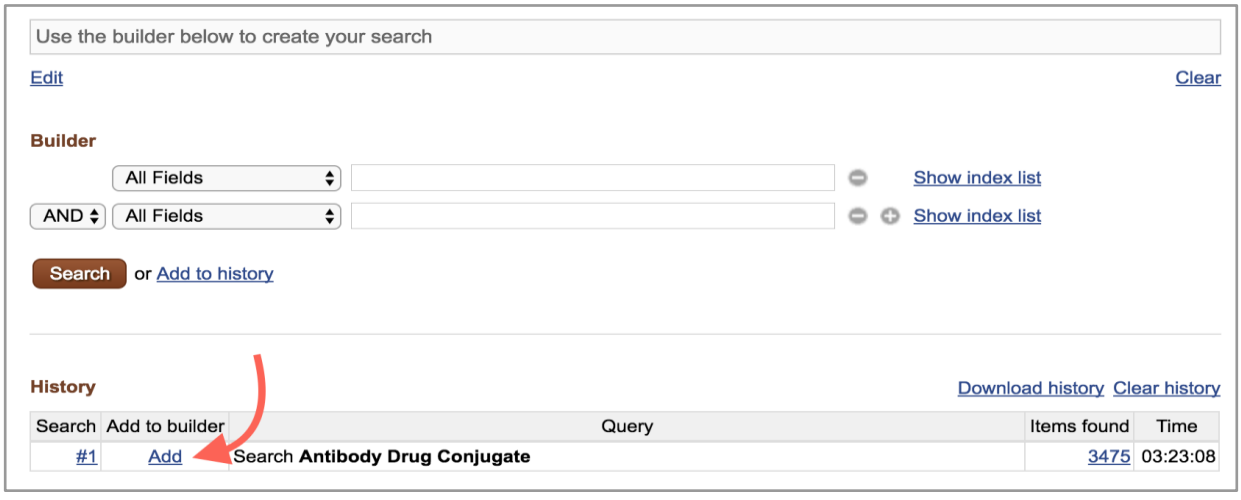
1. Perform a PubMed search using the keyword you want to turn into an alert



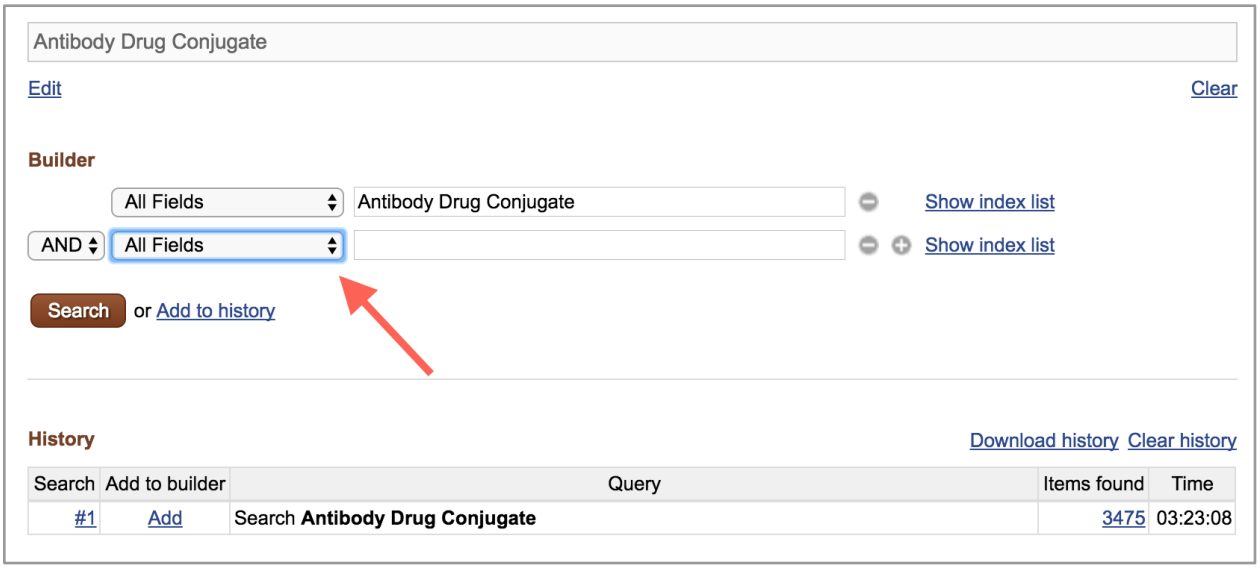
2. Add more restraints to your search query by using advanced search

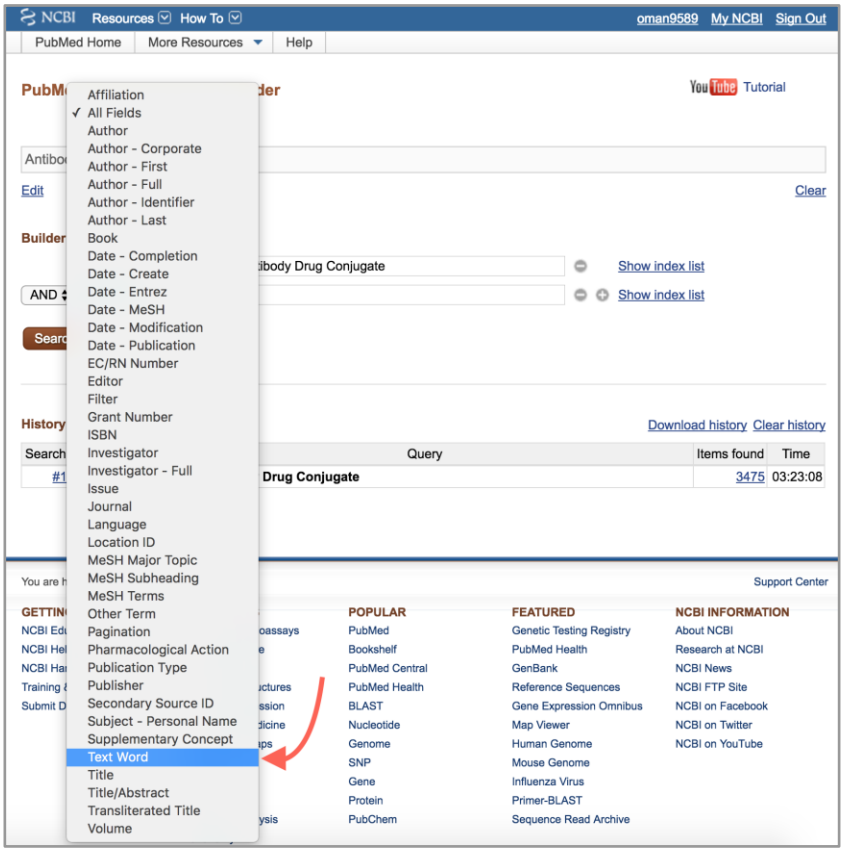


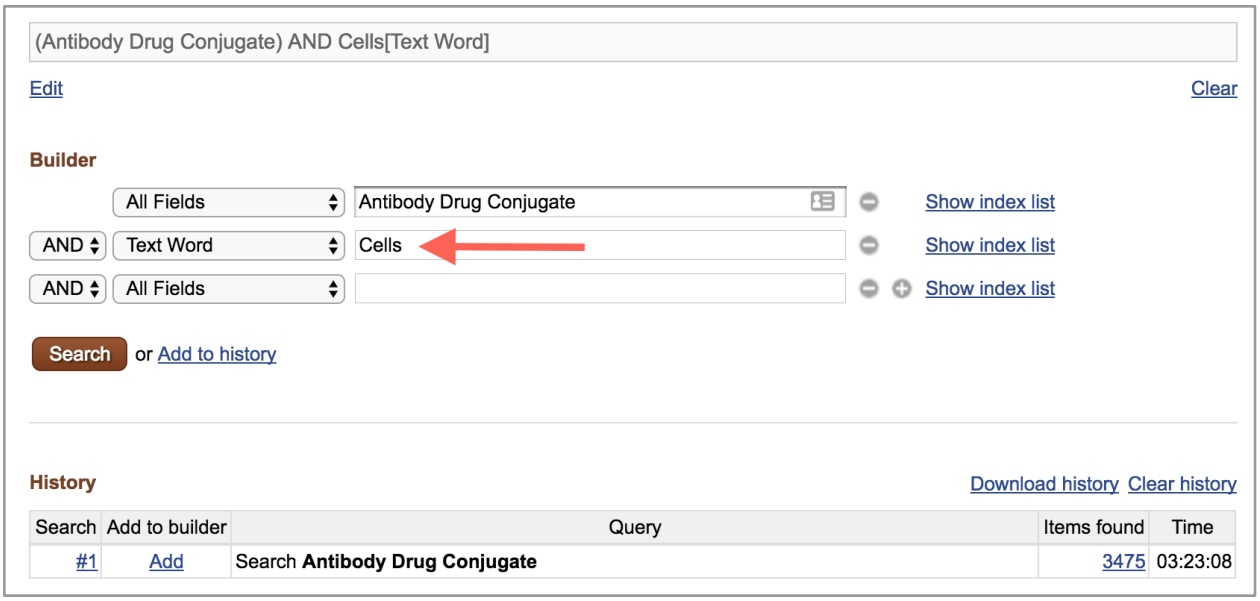
3. After opening the Advanced Search, re-add your original search query



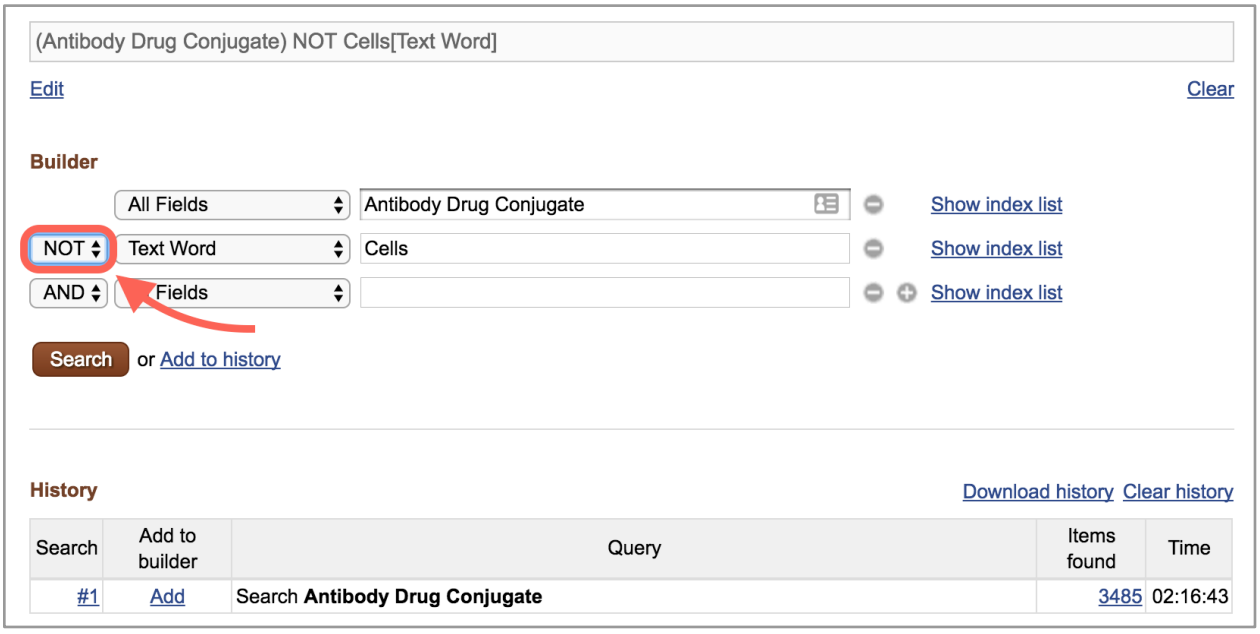
4. Second, add any additional restraints - for this example we will add a text word restraint containing ‘Cells’

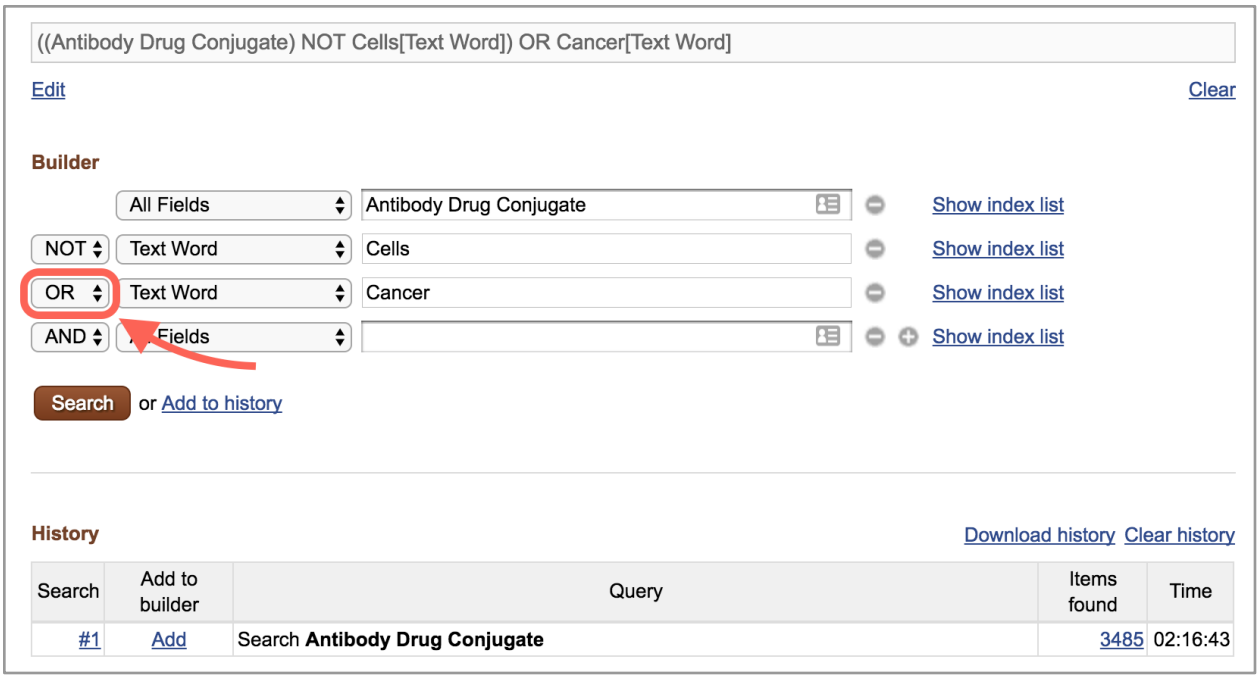




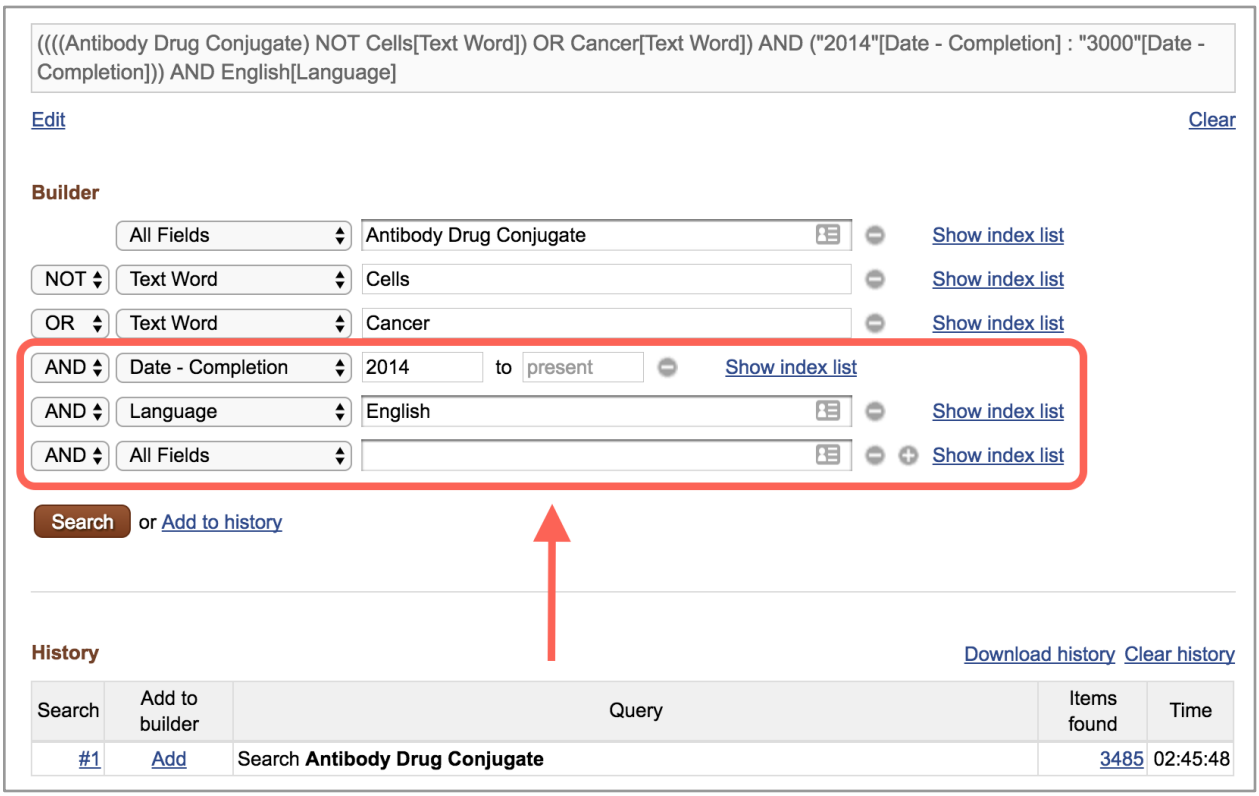


5. To filter a search further, switch the 'AND' to either 'OR' or 'NOT'

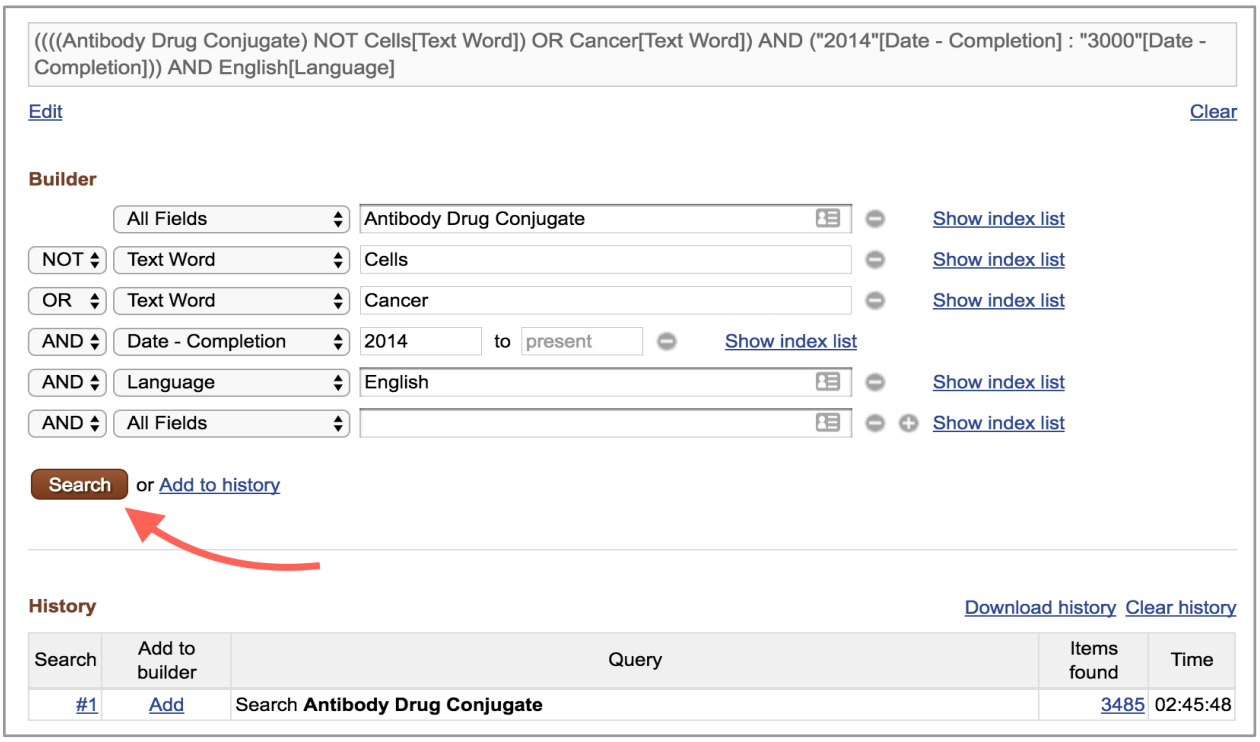




6. Continue to add constraints until your query is complete

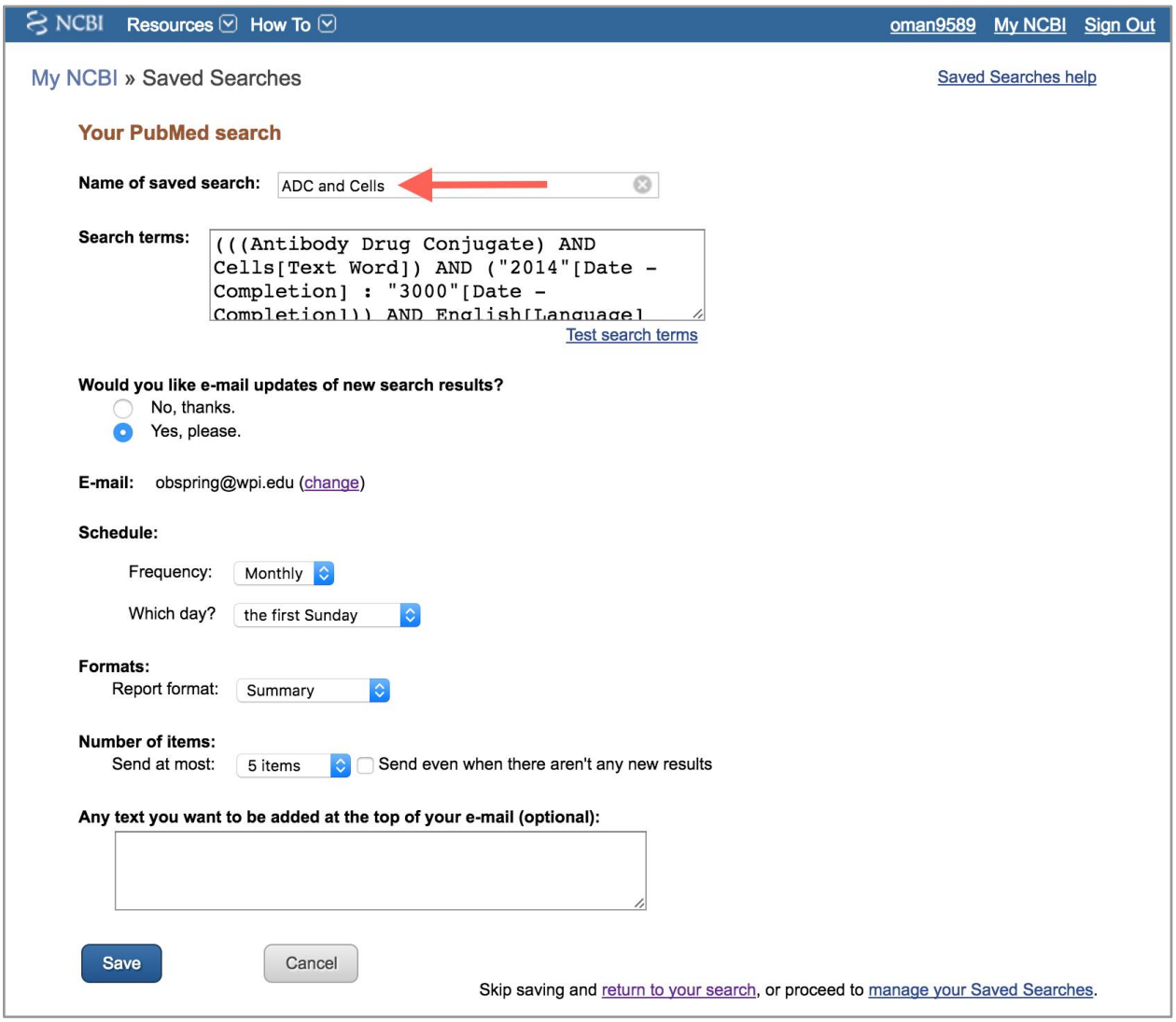


6. Perform your search!

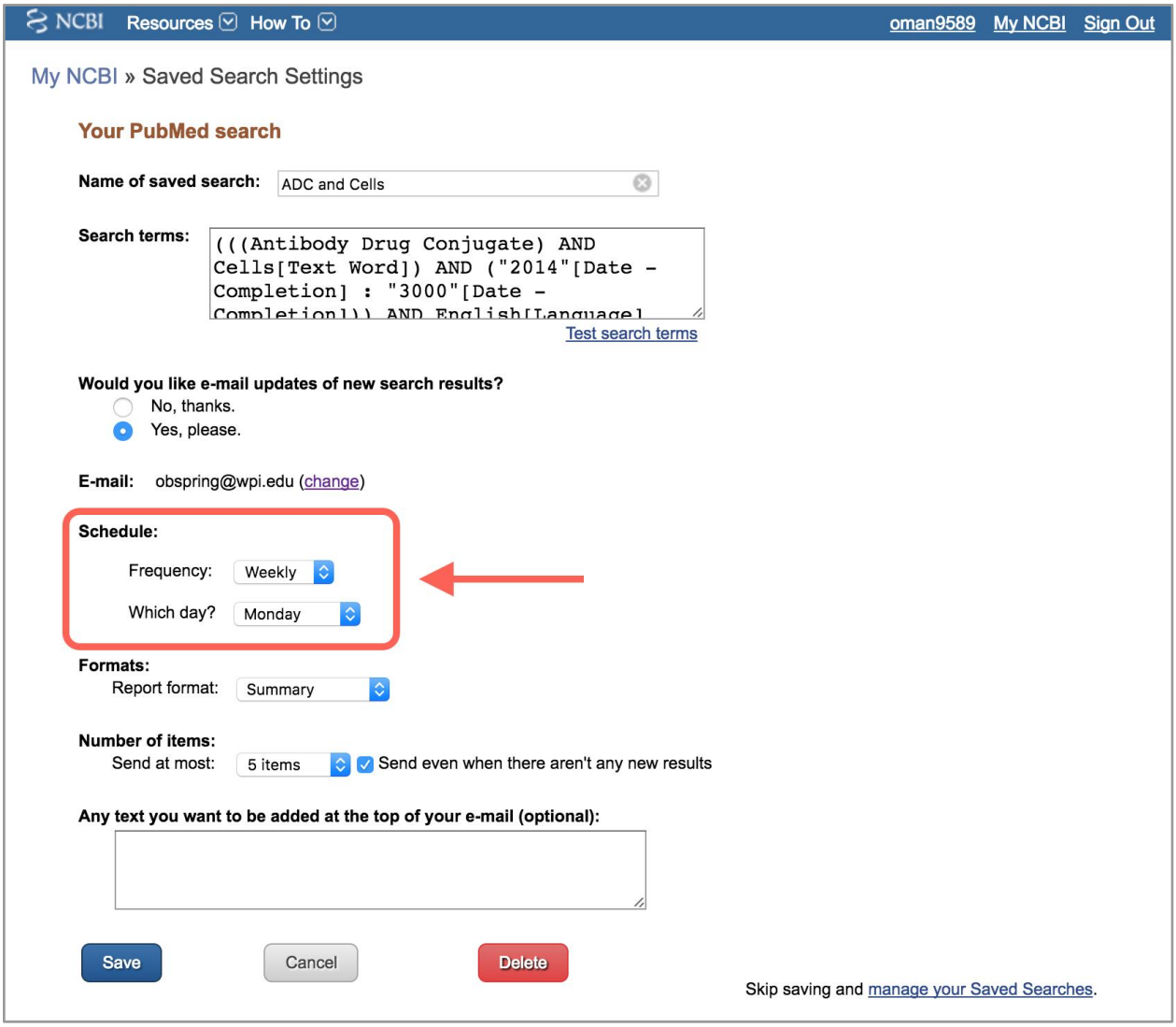


7. Now add your search to your saved searches - click ‘Create Alert’

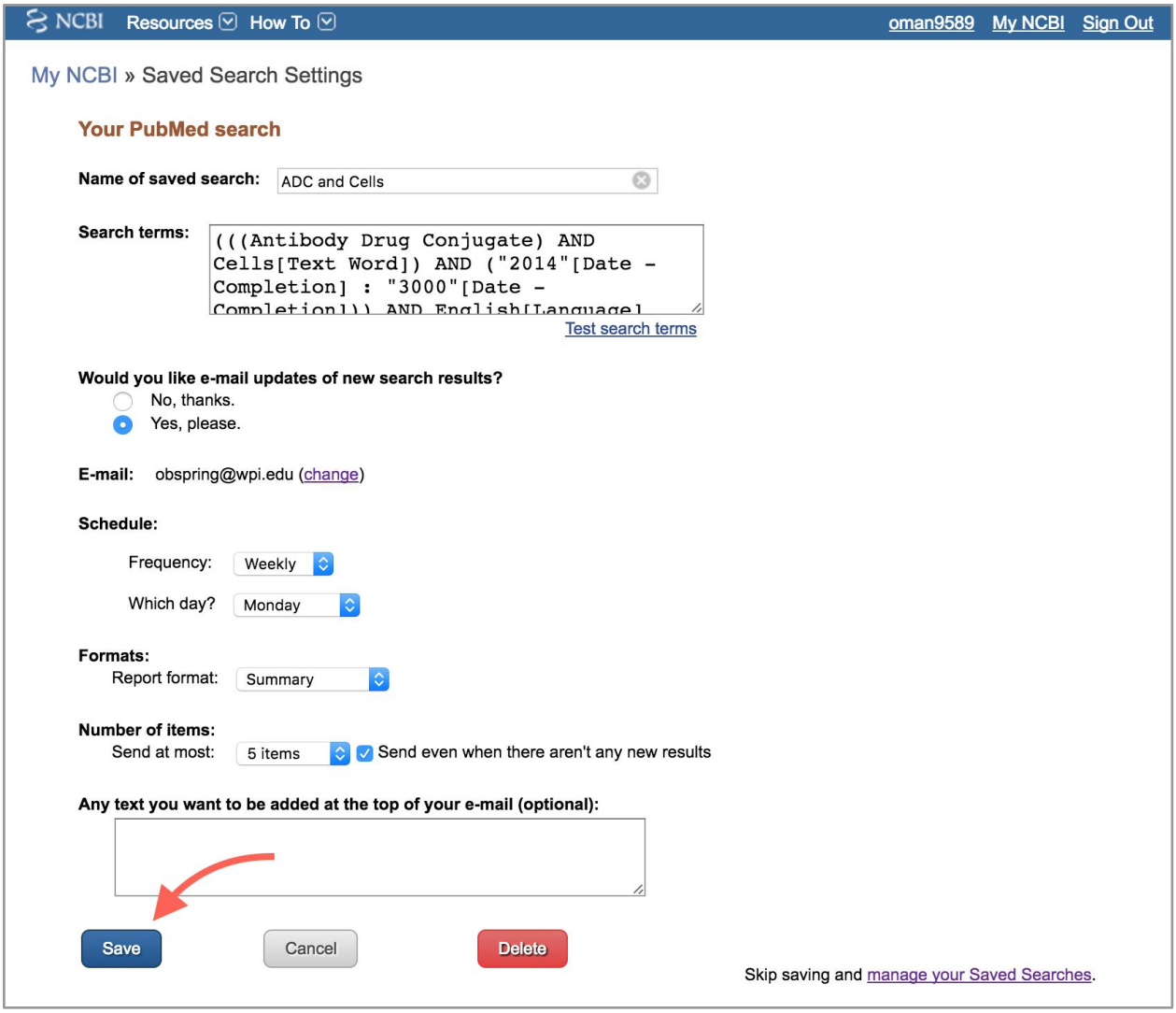
8. Let's name this search 'ADC and Cells' by using the 'Name of saved search' so we can find it later

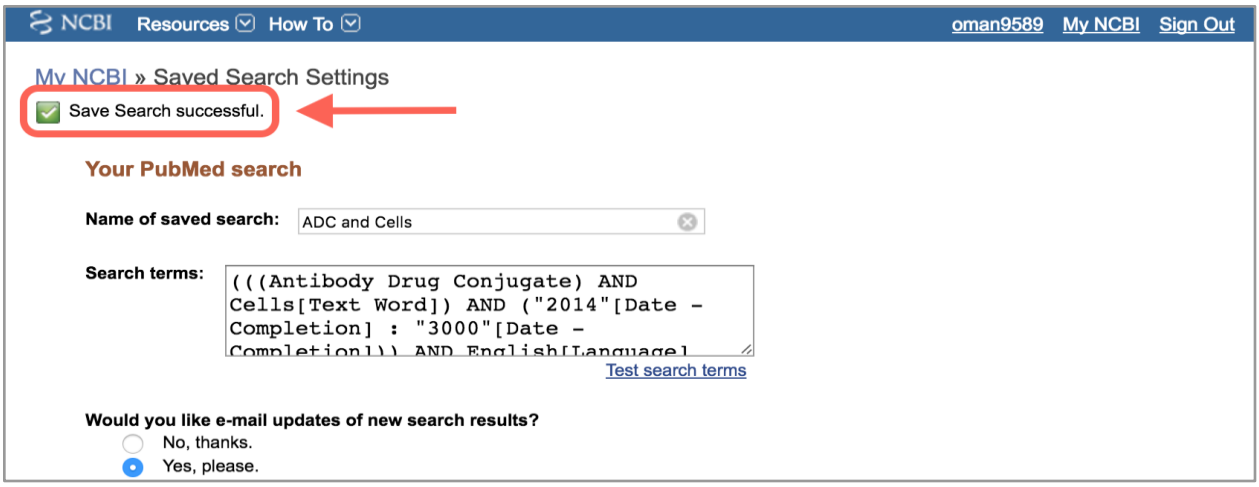


9. Define how frequently you would like the notifications

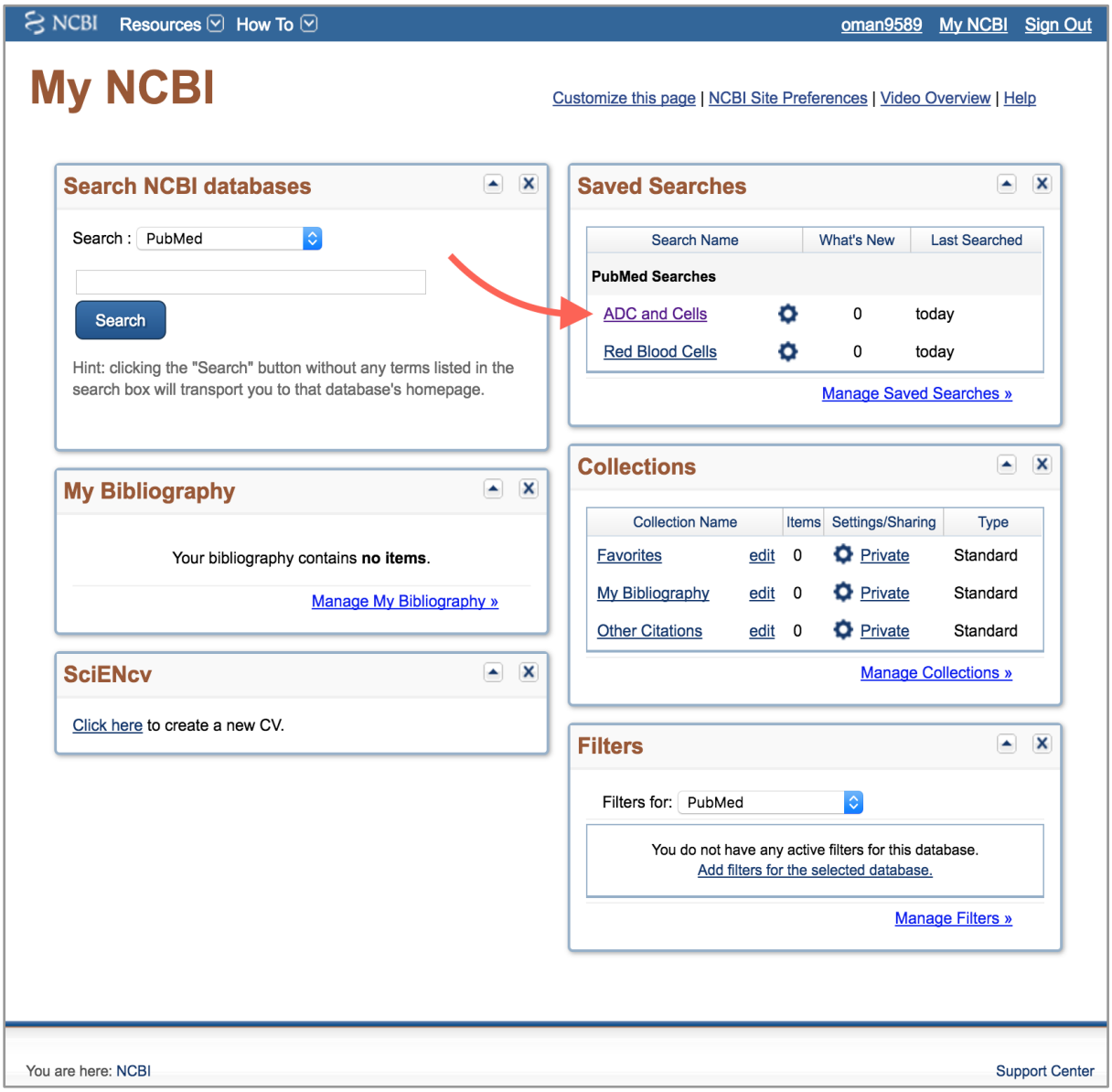


10. Now save your search!



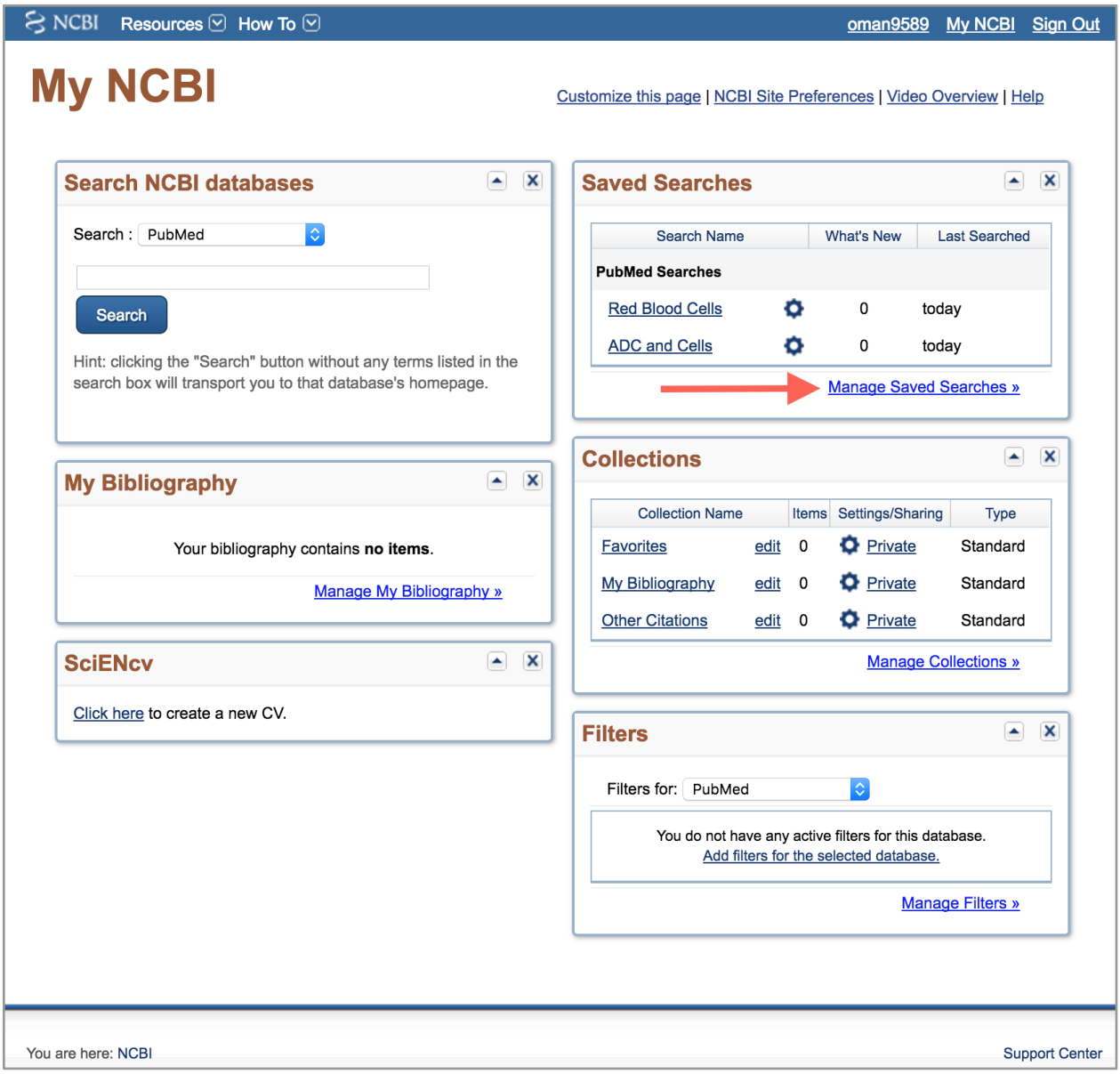

NCBI
Resources
How To
oman9589
My NCBI
Sign Out
My NCBI » Saved Search Settings
Save Search successful.
Your PubMed search
Name of saved search:
ADC and Cells
Search terms:
(((Antibody Drug Conjugate) AND Cells[Text Word]) AND ("2014"[Date - Completion] : "3000"[Date -
Test search terms
Would you like e-mail updates of new search results?
No, thanks.
Yes, please.

11. Now when you navigate back to your ‘My NCBI’ page, you will see your search

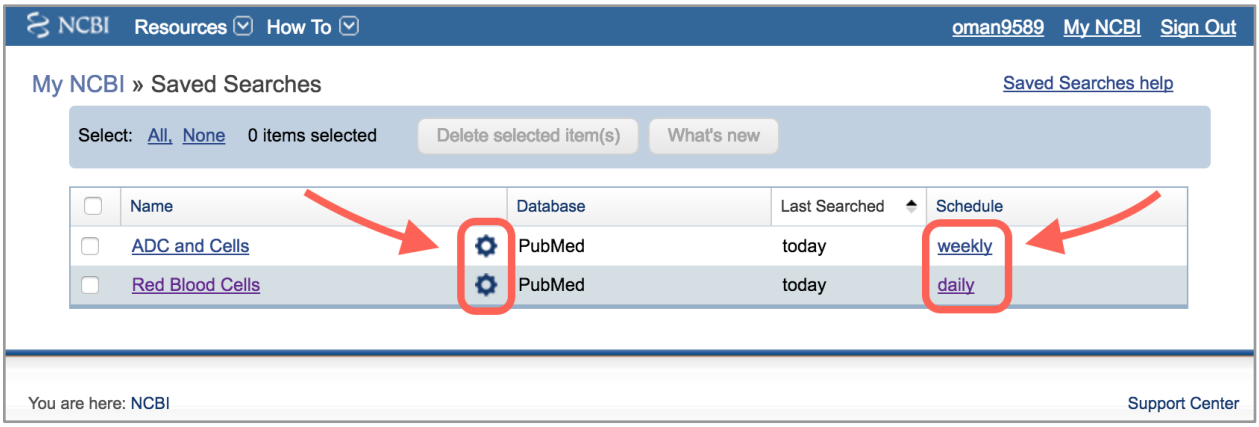


Step 4: Managing and editing saved searches

1. Navigate to the Saved Search management page

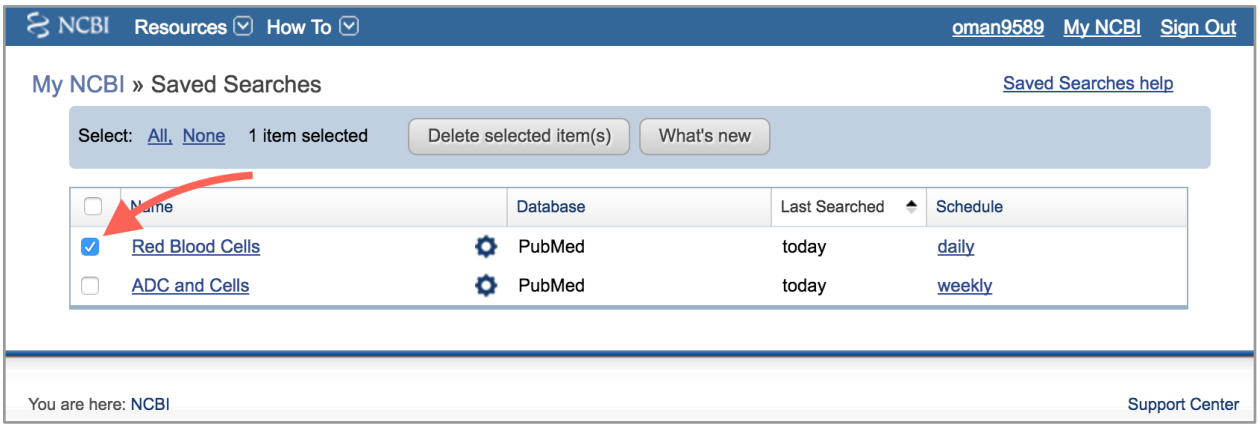


2. In order to edit an alert, click on the gear icon or schedule frequency next to the one you want to edit. They both lead to the same page which is detailed in Step 3 Part 7

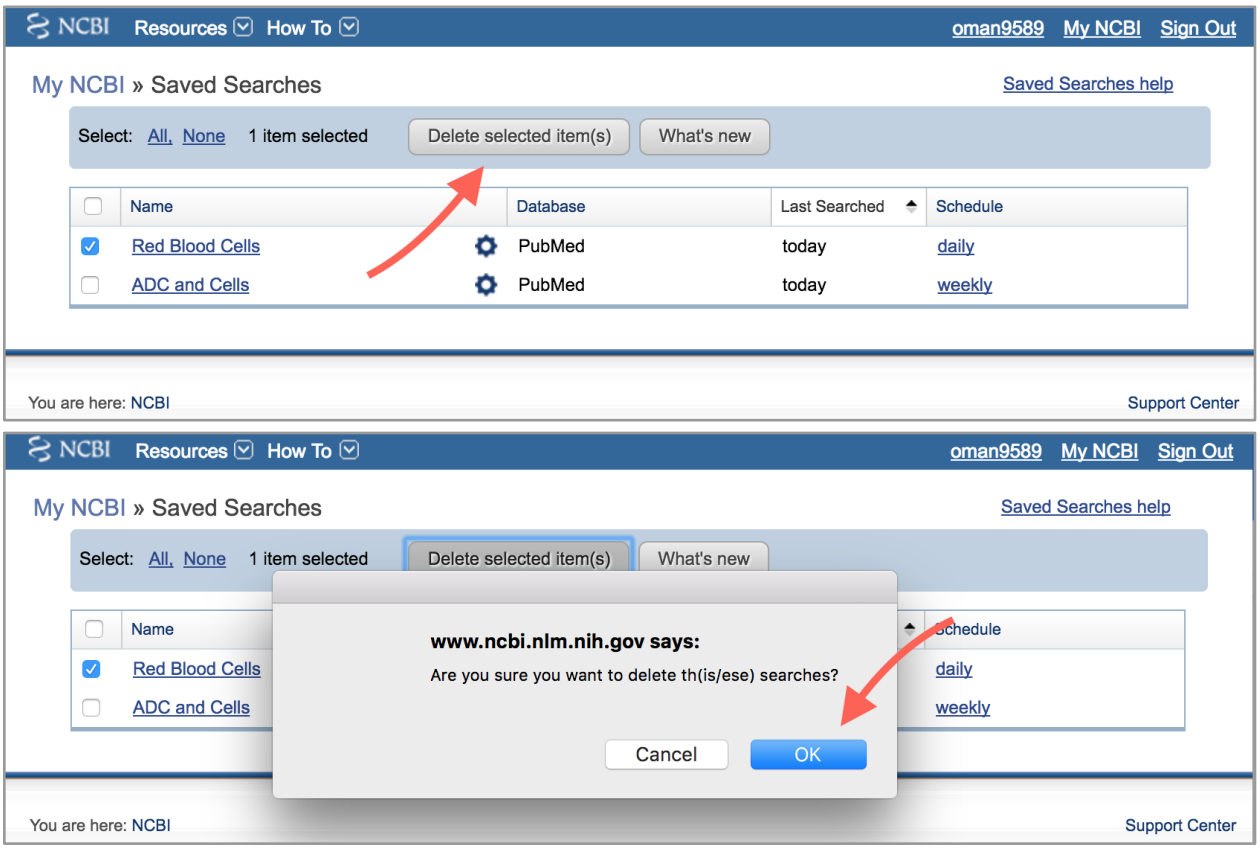


Step 5: Deleting searches

1. After navigating to the Saved Search management page, select the searches you want to delete



2. Click ‘Delete selected item(s)’ and click OK



PAAS Notification Tutorial

Guide to Setting Up U.S. Patent Alerts though the Patent Application Alert Service

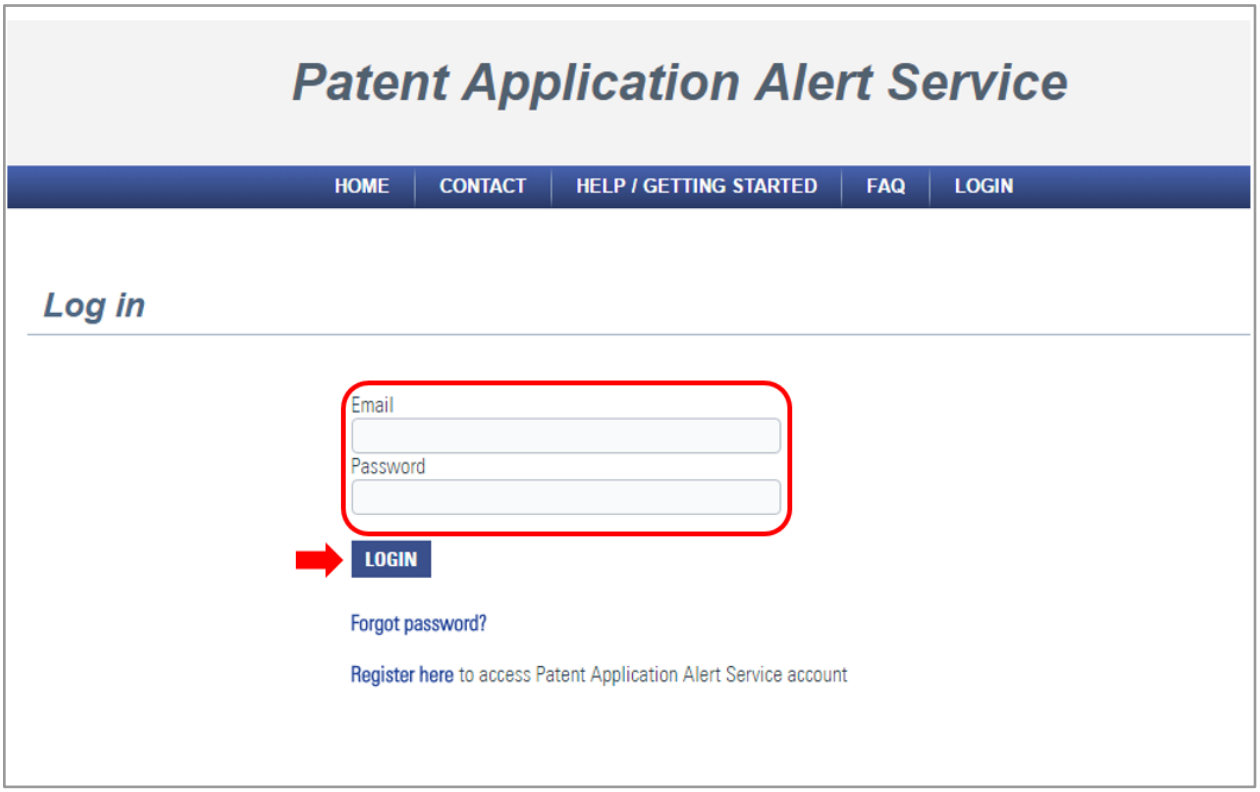
Prepared by Database Team

Step 1: Navigate to PAAS's website - https://www.uspatentappalerts.com/

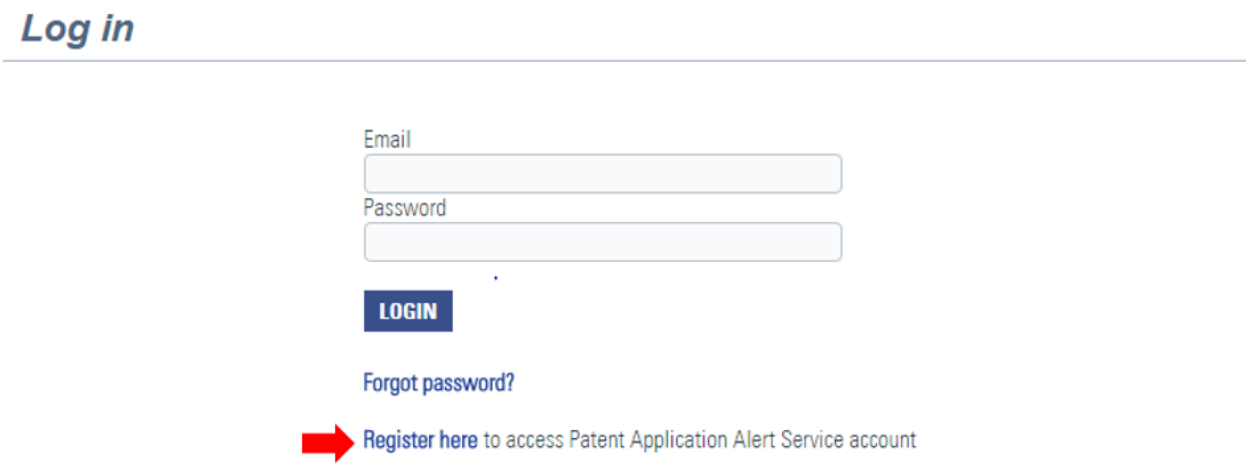
Step 2: Login to your account / Create an account

1. Click Login (Skip to step 3 if already logged in)

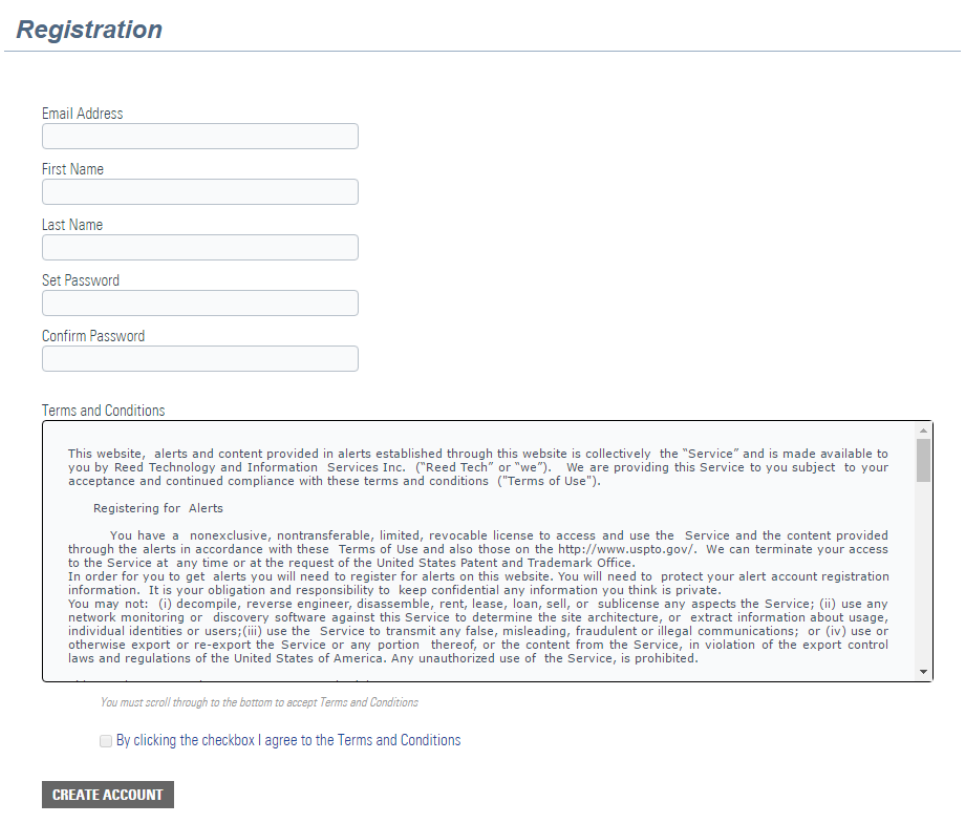
2. If you already have an account, enter your username and password. If not, you need to create an account.



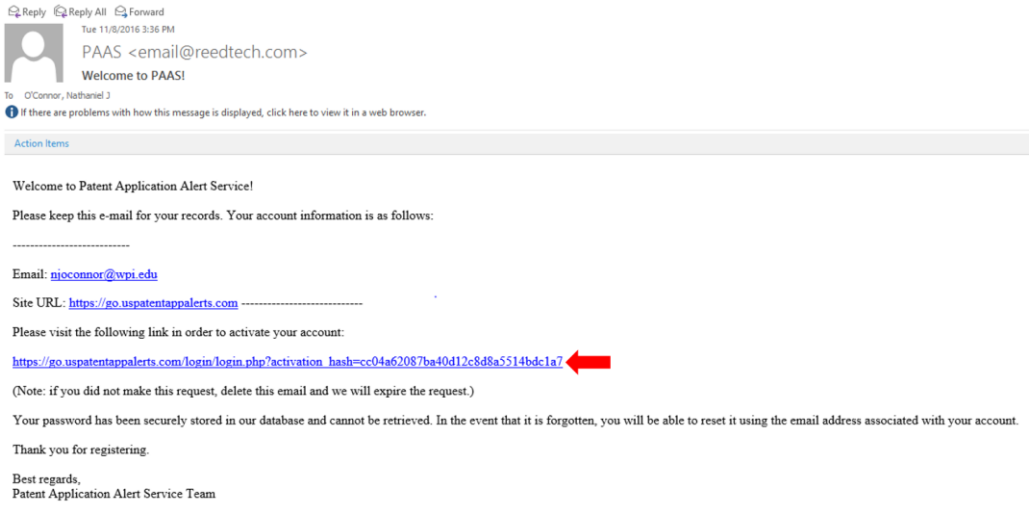
3. To create an account, click ‘Register here’.



4. Fill in your email address, name, and password. Then read the Terms and Conditions and create your account.

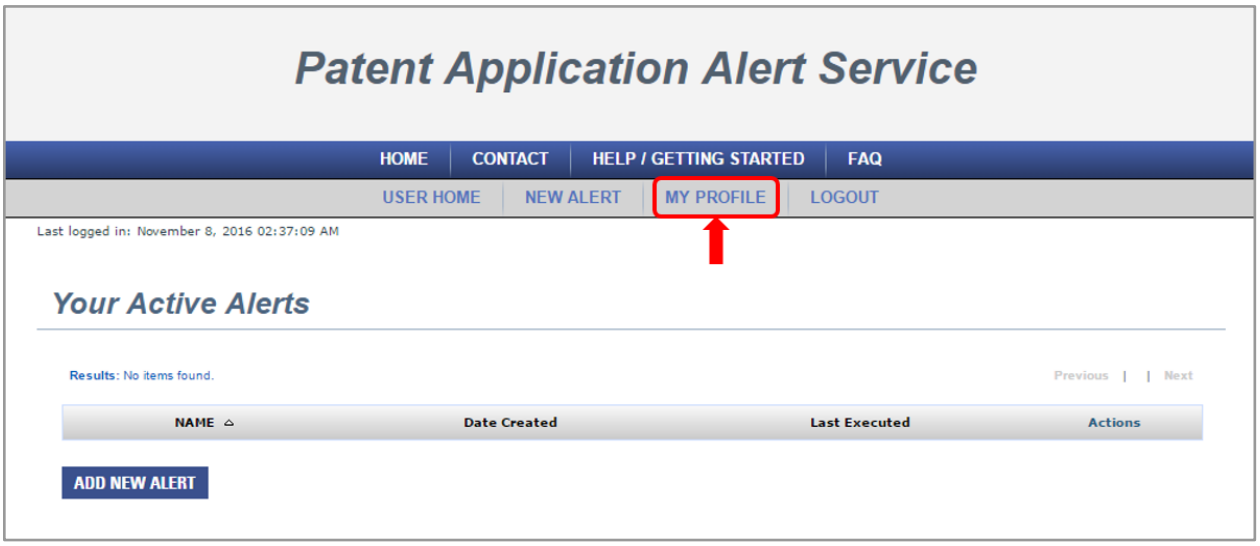


5. You should receive an activation email. Click the bottom link to activate your account and begin creating notifications.

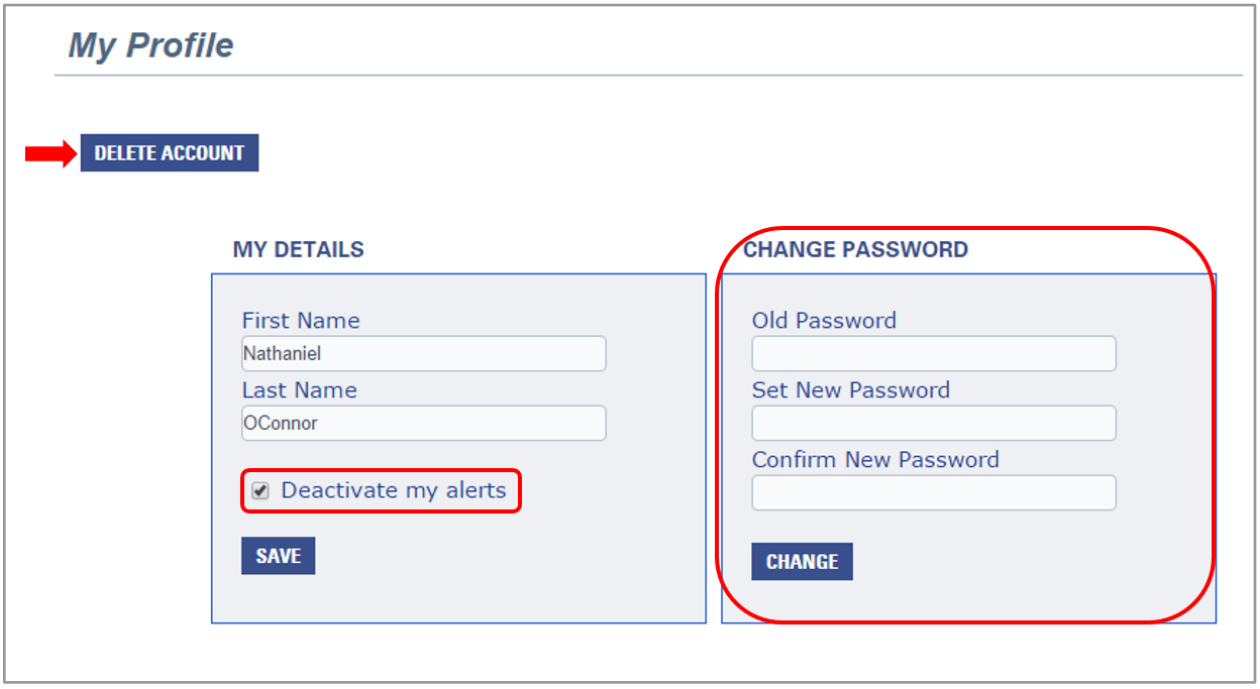


Step 3: Managing your profile

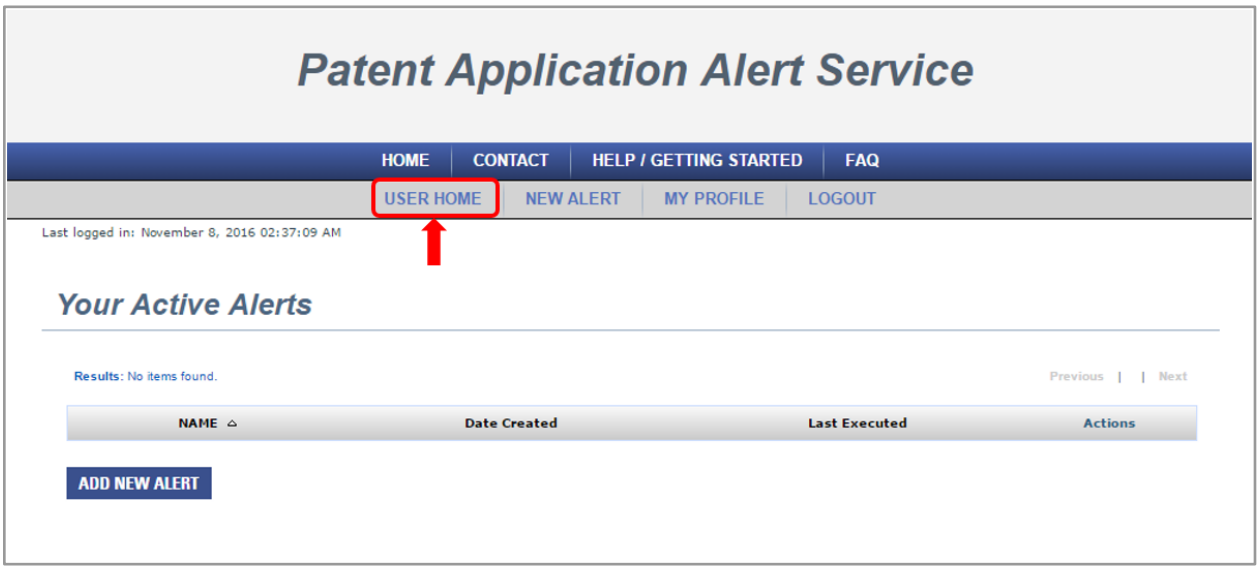
1. Click ‘My Profile’



2. From the ‘My Profile’ page, you can change your password, deactivate your alerts, or delete your account entirely.

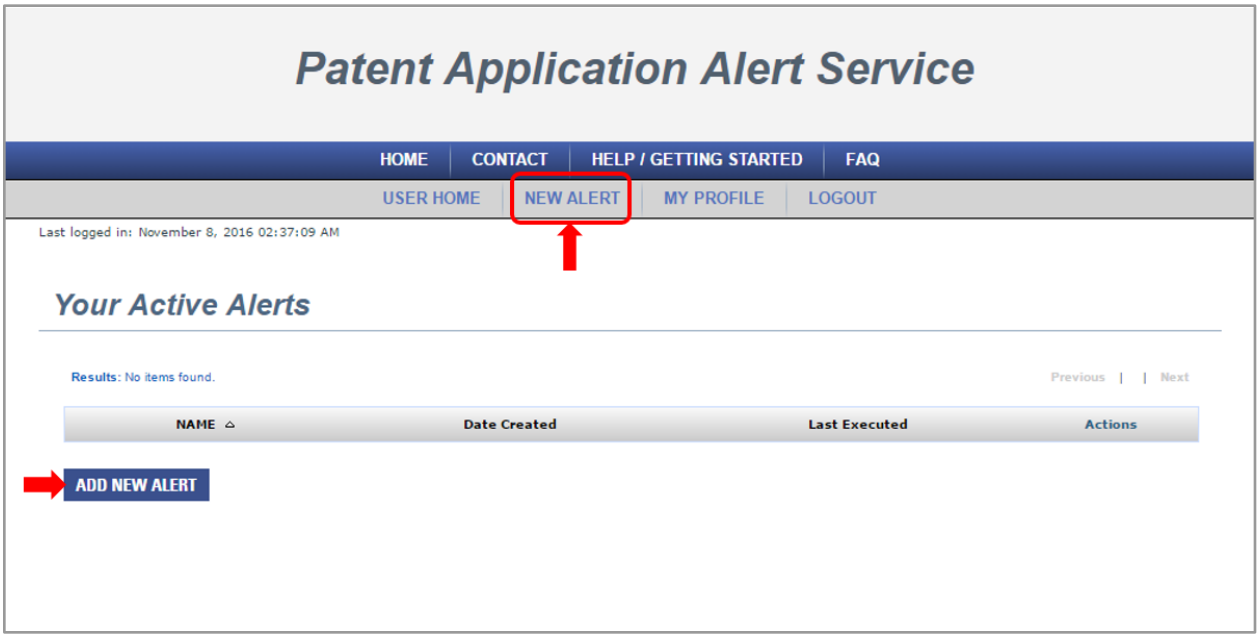


3. Return to the main page at any time by clicking ‘User Home’ on the top ribbon

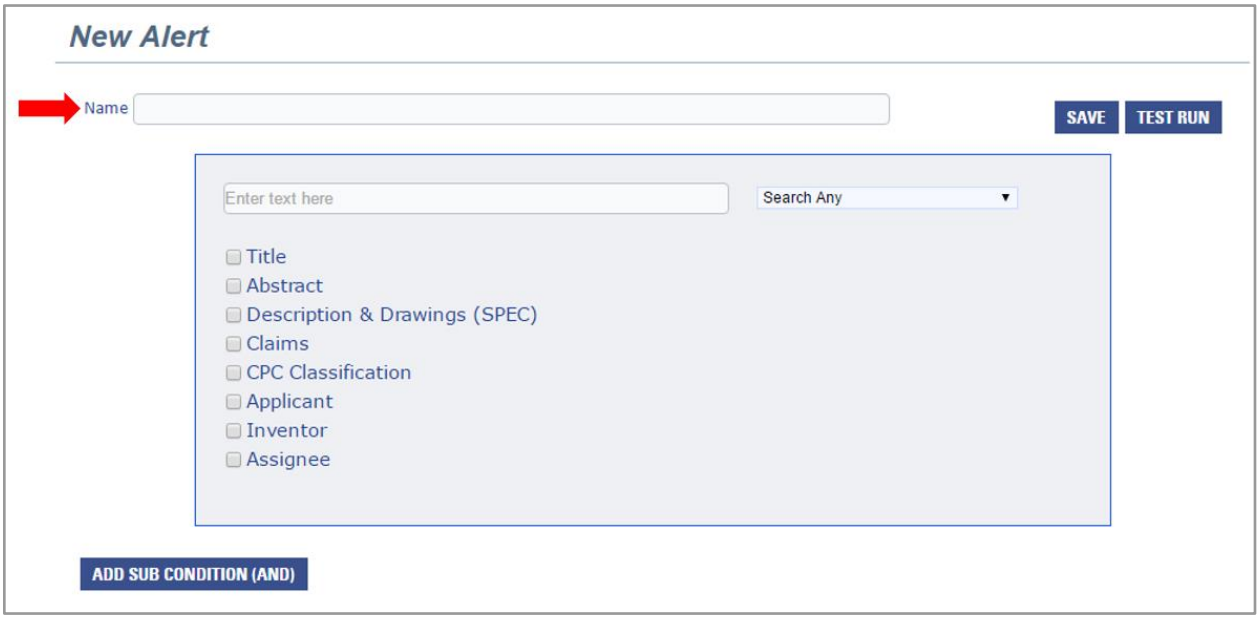


Step 4: Adding New Alerts

1. Begin by clicking either ‘New Alert” in the top ribbon, or ‘Add New Alert’ under your active alerts.



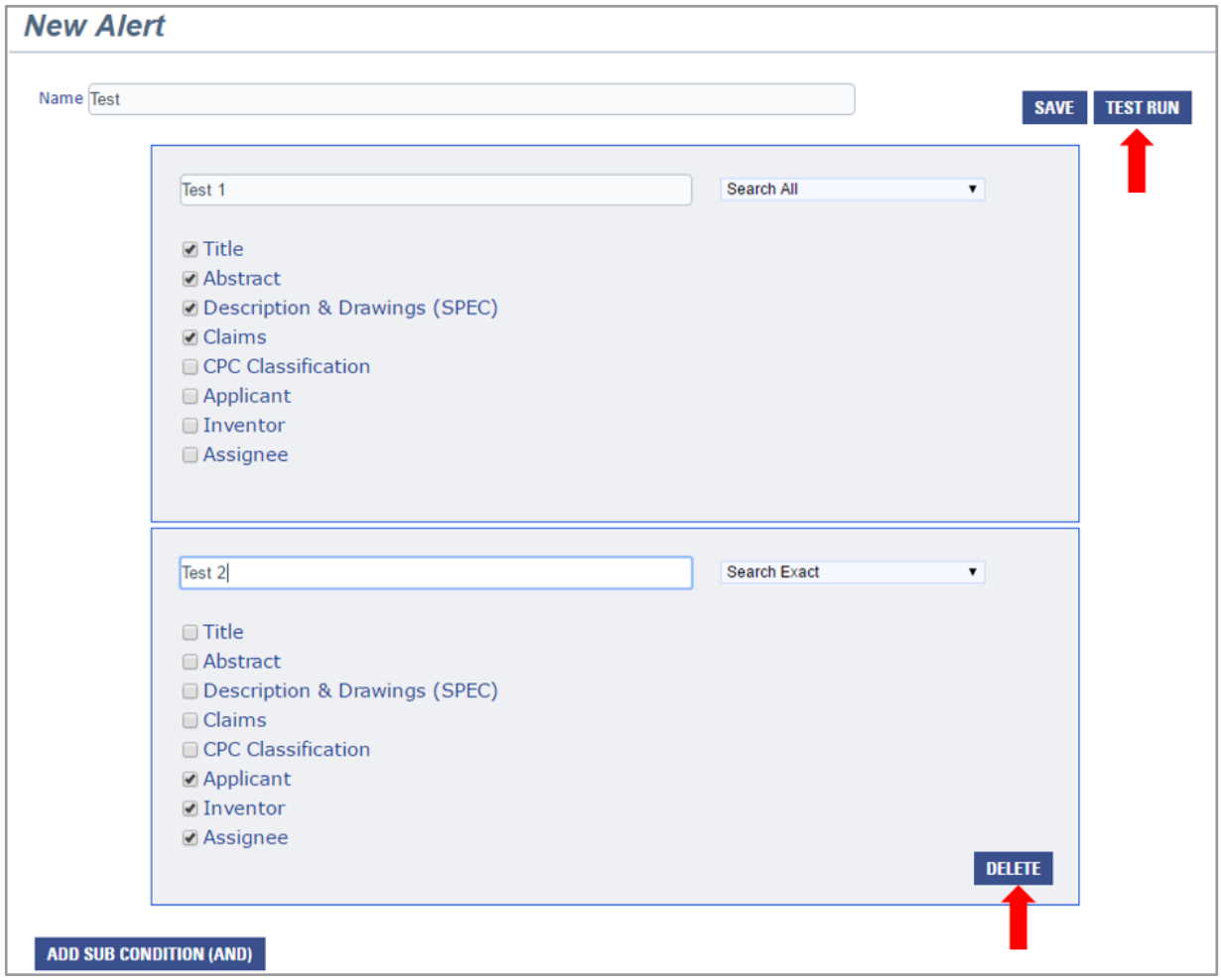
2. Start by naming your alert in the ‘Name’ box. Naming your alerts allows you to manage them later. The name will also appear in your notification email.



3. Inside the box is where you add search conditions. Write the keywords you want to search for in the ‘Enter text here’ box. You can check off any number of boxes as fields to search in. In the top right, the drop down box has 3 options: ‘Search Any’, ‘Search All’, and ‘Search Exact’. ‘Search Any’ will return results containing any of the keywords. ‘Search All’ will return results containing all of the keywords used. ‘Search Exact’ will return results containing all of the keywords in the exact order you wrote them.

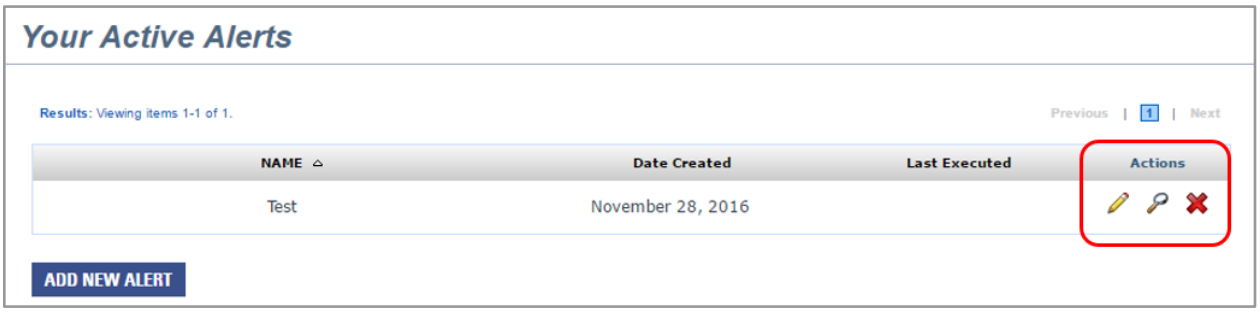
4. If you wanted to search for some words in one field and different words in another field, you can add a sub condition. The sub conditions function like the Boolean operator “AND”.

5. Once you are satisfied with your conditions, click ‘Test Run’ to return results from last week’s publications. You can use this to check if you are getting the results you intended. You can then return to the conditions to modify them if necessary, or click ‘Save’ to save your search and begin receiving weekly notifications. To delete a sub condition, click ‘Delete’ in the box of that sub condition



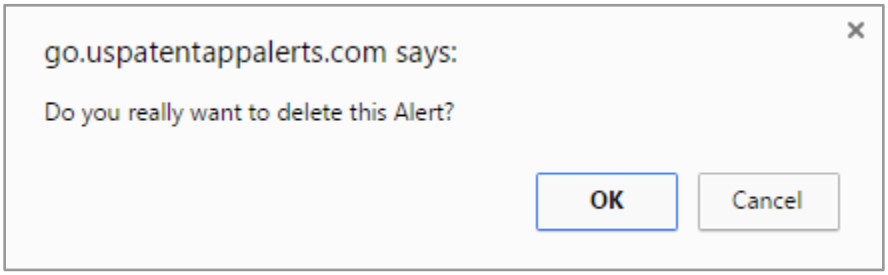
Step 5: Managing your Alerts

1. The home page will display your active alerts.



2. To edit, test, or delete your alerts, use the icons located under ‘Actions’.

3. The pencil icon ‘Edit’ will let you modify the conditions of the alert. The magnifying glass icon ‘Test’ will do a test run of your conditions on the previous week’s patent publications. The red X ‘Delete’ icon will delete the alert.



Note that the USPTO publishes patents on a weekly basis. Therefore notifications are only able to be emailed once per week, at the time of publication.

Appendix O: Factor Calculations

In this section, we go into detail about the calculations utilized to determine the scores for each database in our decision matrix. We utilize a skewed weighting system suggested by Gary Pollice, Professor of Computer Science at Worcester Polytechnic Institute. Our sponsor rated each factor as either very important, somewhat important, or not important. These factors receive a weight of 10, 3, and 1 respectively. We scaled the total weight to 1 by dividing each weight by the total weight. The weights for each factor are shown in Table 9.

Table 10. Factor weights

Factor	Weights x/10	Scaled Weights
Cost	10.00	0.27
Search Result Relevance	10.00	0.27
Advanced Search Options	3.00	0.08
Need for VPN	10.00	0.27
Number of Articles	3.00	0.08
Notification System	1.00	0.03
Total Weights:	37.00	1.00

We calculate the score for each factor differently and scale the results between 0 and 5. We do this to calculate a final weighted average of all factors between 0 and 5 for each literature database.

Cost

To generate a score for cost, we evaluate the cost per article of each literature databases. ScienceDirect and DeepDyve are subscription-based databases which have a flat yearly cost. NCBI and Google require users to pay per article from the publishers. PLOS and DOAJ are both open access, so all the articles found on these literature databases are free. We assume that each researcher will require 1 article per month. Hangzhou DAC Biotech currently has 30 researchers, so over an entire year the company would need to purchase 360 articles. Based on communication with our sponsor, the average cost for their articles was estimated at 40 USD.

PLOS and DOAJ are free and therefore should receive the best score for cost. Their cost per article would be 0 USD. NCBI and Google always require paying per article for articles from publishers, so their cost per article would be 40 USD. ScienceDirect and DeepDyve's cost are calculated by taking their yearly subscription cost, as estimated by the sales rep from each company, and dividing by the 360 articles required for one year. ScienceDirect's yearly subscription cost was estimated at 42,360 USD, so their cost per article is 117.67 USD. DeepDyve's yearly subscription cost was estimated at 10,200 USD, so their cost per article is 28.33 USD.

Using the cost per article from each database, we calculate a score utilizing the following formula:

$$S = 5 \cdot \left(1 - \frac{x}{C}\right)$$

Where,
S = Score
x = Cost per article of each literature database
C = Cost per article of most expensive database

Utilizing this formula, ScienceDirect receives a score of 0 because it is the most expensive option we evaluate. PLOS and DOAJ receive a score of 5 because they are the least expensive. This formula then scales the remaining 3 literature database costs per article between these two extremes. See Table 10.

Table 11. Cost scores

	NCBI	ScienceDirect	Google	DeepDyve	PLOS	DOAJ
Cost per Article	40 USD	117.67 USD	40 USD	28.33 USD	0 USD	0 USD
Score	3.30	0.00	3.30	3.80	5.00	5.00

Relevance

For our relevance test, we weigh the first search results to be worth more than later results. Stephan Sturm, Professor of Mathematics at Worcester Polytechnic Institute, suggested we utilize an exponential weighting for this test. We use the geometric sequence $\left(\frac{3}{4}\right)^{k}$ where k equals the number of the result. The weights of the first five results in descending order are 0.7500, 0.5625, 0.4219, 0.3164, and 0.2373. The total weight is normalized to 1, making the normalized weights 0.3278, 0.2458, 0.1844, 0.1383, 0.1037. Researchers evaluate each search result as either not relevant, somewhat relevant, or relevant. Each keyword for each database receives a relevance score of 0, 0.5, and 1 respectively.

We collect keywords by department, such that researchers evaluate the relevance of topics they are familiar with. We collect these keywords through a survey, which yields an unequal number of keywords for each department. We average the scores provided by each researcher and weigh these averages based on the number of keywords they evaluated. Each department has a relevance score result, which is weighted by the total number of keywords evaluated by that department. We average all the weighted keyword scores for a particular database, giving an average search result relevance score which can be used for comparison. A total of 37 keywords were evaluated; 19 from biology, 7 from chemistry, and 11 from pharmacology. All of the data is in Appendix M. Table 2 in Chapter 4 contains the results of the relevance tests.

Number of Articles

We calculate a score for number of articles by comparing all the literature databases to the largest and smallest of those evaluated. The largest database receives a 5, the smallest receives a 0, and the rest are scaled accordingly.

We calculate a score utilizing the following formula:

$$S = 5 \cdot \left(\frac{x-n}{N}\right)$$

Where,
S = Score
x = Number of articles in a literature database
n = Number of articles in smallest literature database
N = Number of articles in largest literature database

Table 12. Number of articles

	NCBI	Science Direct	Google	DeepDyve	PLOS	DOAJ
Factor	Score	Score	Score	Score	Score	Score
Number of articles:	26,000,000	13,000,000	160,000,000	12,000,000	160,000	2,000,000
Total Weighted Score:	0.162	0.080	1.000	0.074	0.000	0.012
Scaled to 5:	0.808	0.402	5.000	0.370	0.000	0.058

Advanced Search Options

To calculate a score for advanced search options, we select three factors to evaluate each literature database. These factors are Boolean operators, number of filters, and number of ways to sort search results. If a database recognizes Boolean operators while searching, the database receives a 1, otherwise it receives a 0. To evaluate number of filters, the total number of filters for a literature database are divided by the largest total number of filters of all the databases evaluated. Therefore, the database with the largest number of filters receives a 1, and the rest are scaled between 0 and 1. We utilize the same method to calculate a score for number of ways to sort search results. The number of sorting options for each database is divided by the largest number of sorting options of all the databases. We sum the score of each literature database, and then scale these scores to 5. A perfect database would receive a score of 3 by scoring a 1 in all three factors, so we multiply each score by $\frac{5}{3}$ to get a score out of 5. Table 3 in Chapter 4 shows the evaluation of advanced search features.

Notification Systems

To calculate a score for notification systems, we select six factors to evaluate each literature database. These factors are general frequency, detailed frequency, format of notifications sent, number of articles per alert, easy search saving, and easy saved search editing. We evaluated whether a literature database possessed each of these features. Having a feature scores a 1, and lacking that feature scores a 0. We sum these scores to get a total score. A perfect literature database would receive a total score of 6. Therefore each literature database score is multiplied by $\frac{5}{6}$ to scale the score to 5. Table 4 from Chapter 4 shows the evaluation of notification systems.