# INLS 201 Spring 2015 Final Exam

## Ryan Shaw

**Put your name in the header.**

There are eight questions on this exam, worth a total of 32 points.

Question #1 4 points

Question #2 6 points

Question #3 4 points

Question #4 3 points

Question #5 3 points

Question #6 3 points

Question #7 4 points

Question #8 5 points

In addition, there is one bonus question worth 2 points.

**You are permitted to consult the readings, my slides, your notes, or anything else you wish, but you must work on the exam alone.**

Use as much space as you need for each answer; the formatting of this exam is not intended to indicate how much you should write.

When you are finished, please email it to [ryanshaw@unc.edu](mailto:ryanshaw@unc.edu) with the subject line “INLS 201 Spring 2015 Final”.

This exam is due by 11AM on Friday, May 1, 2015.

## Question #1 (4 points)

In her chapter on “Models of the Information Seeking Process,” Hearst quotes Wilson (1981),

[W]hen we talk of users' “information needs” we should not have in mind some conception of a fundamental, innate, cognitive or emotional “need” for information, but a conception of information (facts, data, opinion, advice) as one means towards the end of satisfying such fundamental needs.

Wilson is arguing that talk about “information needs” is somewhat misleading, because it confuses means and ends. When one seeks information, it is not because *information* is needed; rather there is *something else* one needs or needs to do, and obtaining information is a means toward that end.

For example, if my child were sick I might use a search engine to look for information about her symptoms because I need to make a decision about how to care for her. What I really need is not information, but to make a decision (and ultimately I need for my child to be healthy). Information is just part of the means to that end.

What are the implications of Wilson's argument for the **standard model of information seeking**? How might we revise the standard model to better reflect Wilson's way of thinking?

The standard model of information seeking begins with a task and is completed with (a) result(s) from a corpus of documents. However, this does not fit with Wilson’s argument about “information needs.” She argues that information needs do not simply end with information from one or more documents out of a corpus, but the information from those documents are used by a person to reach means ends. If we were to revise the standard model to take this into consideration, it would, after acquiring information from one or more documents, evaluate to determine whether or not it is applicable to reach the solution to the task at hand. Then, externally of the information seeking process, is the information applied to reach and end of the task at hand. The largest issue of this however, is that this is altering the information seeking process to an information acquisition and execution process, merging former information seeking process with Norman’s cognitive execution and evaluation process.

The revised standard model is as below:

## Question #2 (2 parts, 6 points total)

Execution

Evaluation

External Application and Evaluation Process

Search Engine

Corpus

Query

Consider the following usability measures:

* Time to learn
* Speed of performance
* Retention over time
* Rate of error
* The North Carolina State Legislature, concerned about fraudulent voting, has passed a law requiring election officials to review every ballot cast in order to ensure its validity. Advocates for college students, minorities, and the elderly are concerned that overly zealous officials will incorrectly invalidate ballots cast by members of these groups. You have been hired to evaluate the software that officials will use to review ballots.  
    
  We can define the **efficiency** of an official using the tool as the **correctness** and **completeness** with which the official reviews ballots, divided by the **amount of resources** (time and effort) expended. Which of the usability measures above do you think will have the **most** impact on efficiency? You may select more than one measure. Explain your reasoning. **(4 points)**

Retention over time would have the greatest impact on efficiency because of its impact on rate of error. With low retention rates, as time moves forward, the rate of error will increase at the same rate that retention is lost. While time to learn the software and speed of performance are static variables over time, rate of error is not and is dependent upon the retention rate over time. With a retention rate lower than 100%, efficiency will decrease. Once correctness and completeness decrease, the only way to regain them would be to spend more resources on retraining the voting officials (thus compounding time to learn the software), resulting in another loss of efficiency. Below is an equation where correctness (t), completeness (t) and resources (t) are functions of time for all instances where retention over time is not 100%.

\*assuming ∞ is the lifetime of an employee.

* You have been hired as a usability consultant by The Common Application to help them evaluate the usability of their interface for applying to college. Which of the usability measures above would you focus on improving, and why? **(2 points)**

Unlike elections, most people apply to colleges at only one point in their life and, for the most part, use the Common Application less than 10 times. This puts a high weight on the rate of error and time to learn (as a user has only a few months to apply). If the resources users spent learning the Common Application outweighed the benefit, schools would drop the Common Application for a competitor or their own application software. Likewise, if the Common Application had a high rate of error, applicants and schools would not trust the software for such an important task. In this case, I would focus on improving the rate of error. A single error could mean the difference between an acceptance and rejection letter. The second highest priority would be eliminating time to learn the software. I put this second to rate of error however because if the time to learn the software is high, the universities can change the amount of time that applicants can work on their application (up to a reasonable amount). They, however, cannot change their application process if there are bugs in the software that allow for an error of communication between the applicant and the school.

## Question #3 (4 points)

According to Shneiderman and Plaisant, one of the key usability measures is **subjective satisfaction**. Think about satisfaction as one measure of the quality of an information retrieval (IR) system. What does it mean to be subjectively satisfied with an IR system? Is this a good measure of quality? Can we be satisfied with things that are not good for us? Are there alternative measures we might consider?

Shneiderman and Plaisant define subjective satisfaction, as how much users liked using various aspects of the interface.

Each information retrieval system has a different goal. Google search’s goal is to have their users more satisfied than those of the competitors – the main reason being profit, thus putting a high value on subjective satisfaction. An IR system within a university’s library however may have the goal to return results that contain factual and peer-reviewed information. Here, subjective satisfaction plays a less dominant role – nonetheless, it is still valued; if a interface is extremely difficult to use or learn, a researcher may choose other means to retrieve the information they are seeking (i.e. reference of a human librarian or specialist in the field they want to learn about).

With the varying goals of information retrieval systems, it is not a good measure of quality. A competitive commercial IR system like Google or Bing need high subjective satisfaction rates to retain their users and their information for profit. The quality and quantity of their search results may play a role in satisfaction, but it is not necessary for the results to be quality. In a research setting however, it is paramount that the retrieved documents are factual and peer-reviewed and user satisfaction may be less important. Whether or not an IR system remains a dominant one among competitors is dependent on many more factors than just subjective satisfaction, and the weight of those factors change with the setting and goals of the IR system.

## Question #4 (3 points)

In their chapter on “Usability of Interactive Systems” Shneiderman and Plaisant claim that the ultimate goal of user interface designers is “universal usability”: addressing the needs of the broadest possible set of users. Achieving universal usability is difficult because it requires understanding the many ways that users and contexts of use can differ.

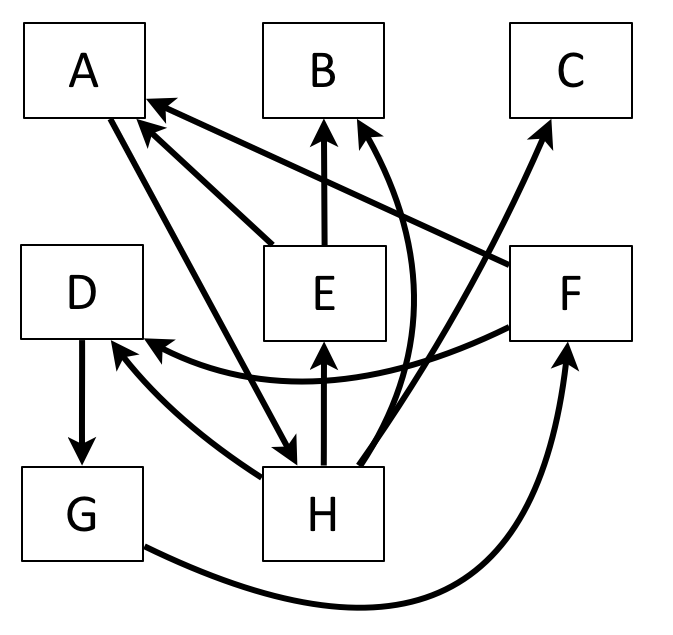
Identify and discuss **three** kinds of differences between users, or between contexts of use, that ought to be taken into consideration when evaluating the usability of **web search interfaces on smartphones**. For each difference that you identify, explain why it is relevant to evaluating usability.

The first major difference between users that needs to be addressed when evaluating the usability of a mobile web search interface is the difference in cognitive and perceptual abilities. A new user to web search may type a query exactly as they think of it, whereas a person who has experience with web search may know some tricks for web search break their query into keywords and may use certain tags in their query. Both users in these instances should be able to have appropriate results, despite their differing background in web search. Likewise, the search engine should take into consideration differences in vocabulary sizes by using synonyms in the query transformation process.

The second major difference between users that needs to be addressed when evaluating the usability of a mobile web search interface is the variation of physical abilities. Essential elements on the user interface (search box, buttons, etc.) must be placed in areas that almost all users can easily access regardless of the size or shape of their hand. A user with small hands shouldn’t have to struggle to hit the search button that a user with a large hand can easily press.

The third major difference between users that needs to be addressed when evaluating the usability of a mobile web search interface is cultural differences. Homographs words with the same spelling that have different meanings from region to region, even if they are in the same language. The search engine must take into consideration the locale of the user. For example, the term “pissed” in the US would return search results around the idea of being angry, whereas in the UK, it would return results around the concept of drunkenness.

## Question #5 (3 points)

Using the graph below, compute the PageRank of **node H** after **two iterations (*k*=2)**. (Note: you do not need to calculate the PageRank of every node in order to answer this question.)

|  |  |  |  |
| --- | --- | --- | --- |
|  | *k*=0 | *k*=1 | *k*=2 |
| A | 1/8 | 1/16 + 1/16 = 1/8 |  |
| B | 1/8 |  |  |
| C | 1/8 |  |  |
| D | 1/8 |  |  |
| E | 1/8 |  |  |
| F | 1/8 |  |  |
| G | 1/8 |  |  |
| H | 1/8 |  | 1/8 |

## Question #6 (3 points)

In “The Design of Search User Interfaces” Hearst identifies the three stages of user-centered design as:

1. Needs assessment
2. Task analysis
3. Design–Evaluate–Re-design

Suppose that you were designing an interface to help UNC students find summer employment, internships, volunteer and study abroad opportunities. Explain what you would do during each of these stages. What activities would you carry out, and what would be the products of each stage?

The first step in the needs assessment, we need to assess the users who will be using them: UNC students, typically between the ages of 18-25 and untrained in the software. Their goal is to find summer employment, internships, or volunteers/study abroad opportunities. The tasks they have to complete to reach this goal are to use the interface to connect to a series of outside employers/opportunities. On the other end of the software, a user (whether it be the employer themselves or a UNC employee acting on behalf of the employer) has to list information about opportunities. The software in the end must satisfy the both the user (by returning many relevant opportunities) and the employer (by effectively organizing the applicants for comparison). At the end of this step, we will have a general idea of what we want our software to do.

The interface on the student end has to be fairly easy to learn and use. On the end that the employer/representative of the employer, the interface would optimally be easy to use, but it wouldn’t be as paramount as the interface on the student end. The student end of the user interface must allow for a student to easily send the same information about themselves to prospective employers (most likely done by an account of some type). The employer end of the software must allow for an introspective analysis of each applicant as well as a way to easily compare applicants (most likely done by a database of some sort). At the end of this step, we should have a general idea for the architecture of the software: a simple to use interface for the UNC students that match them with potential opportunities that a more skilled user uploads.

After we assess the needs and analyze the tasks that need to be done, we will design the UI. After this, we will have users test out the interface, return feedback, and that feedback will shape the next version of the interface. This cycle will repeat until we decide that it is sufficient and meets expectations. At the end of this stage, we should have publishable software for UNC students to use.

## Question #7 (2 parts, 4 points total)

Hearst identifies a number of best practices for designing search user interfaces, including:

* **Show search results immediately**
* Pay attention to interface details
* **Highlight query terms**
* Provide access to search history
* **Integrate navigation and search**
* Avoid empty result sets
* Show informative document surrogates
* **Allow re-sorting of results**
* Support uses with varying vocabularies
* Show query term suggestions
* Support rapid response
* Rank search results
* Transform queries
* Recognize the importance of interface aesthetics
* Provide shortcuts
* Suppose you were designing the interface to an app that enables users to search all the text messages they have ever sent or received. Which of these best practices would you employ to **efficiently inform users about the status of the app**, so that they can easily understand what effects on the app their interactions will have? You may choose more than one best practice.   
    
  For each best practice you choose, explain how you would do this for the specific case of a text message search app. For example, if the best practice were “provide shortcuts,” you might explain that you would do this by putting the names of the 2–3 people with whom the user texts the most at the top of the interface, and when the user taps one of these names, the most recent text messages exchanged with that user would be shown. **(2 points)**

My interface would show the users the status of the app by immediately show search results by incrementally searching as the user typed their query. It would print out the results as it found them, with a highlighted query, instead of waiting for the entire collection to be traversed, and the printed list of results and highlighted text would change as the query changed. It would also list the number of search results (which would also change as the query changed).

* Which of these best practices might you employ to **reduce the amount of mental work users are required to do** when using the app? You may choose more than one best practice. Again, for each best practice you choose, explain how you would do this for the specific case of a text message search app. **(2 points)**

To reduce mental workload, I would have the returned sets highlight the query within the overall text message. I would also allow the user to navigate through the search results with simple “next” and “previous” buttons, easily traversing the corpus of text messages to ones that match the query. The search results could be resorted by different fields such as date sent/received, conversation, sender, or message type (i.e. MMS, SMS, etc.).

## Question #8 (2 parts, 5 points total)

Read the following passage from Nicholas Carr's post, “The searchers” (Rough Type, Jan. 13 2013).

*Ray Kurzweil, the inventor and AI speculator, recently joined [Google] as its director of research. His general focus will be on machine learning and natural language processing. But his particular concern, as he said in a recent interview, will entail reconfiguring the company's search engine to focus not outwardly on the world but inwardly on the user:*

“I envision some years from now that the majority of search queries will be answered without you actually asking. It'll just know this is something that you're going to want to see.” While it may take some years to develop this technology, Kurzweil added that he personally thinks it will be embedded into what Google offers currently, rather than as a stand-alone product necessarily.

*This has actually been Google's great aspiration for a while now. We've already begun to see its consequences in the customized search results the company serves up by tracking and analyzing our behavior. But such “personalization” is only the start. Back in 2006, Eric Schmidt, then the company's CEO, said that Google's “ultimate product” would be a service that would “tell me what I should be typing.” It would give you an answer before you asked a question, obviating the need for searching entirely. This service is beginning to take shape, at least embryonically, in the form of Google Now, which delivers useful information, through your smartphone, before you ask for it. Kurzweil's brief is to accelerate the development of personalized, preemptive information delivery: search without searching.*

* What policy or ethical questions does “search without searching” raise? **(2 points)**

To ever possibly be able to offer a service that “searches without searching,” Google (or any other company) must accumulate vast amounts of data on each of their users and a vast amount of resources to fund the high computational task of providing such a service. One ethical question is what to do in the event that Google predicts something you want to search incorrectly. While one mistake could lead to embarrassment of the user, another could lead to biased or wrong information.

One of the larger ethical questions with this is the selling of your data to fund the computational tasks. Who is buying your data and what is being done to it is out of your hands. The things that are being done with your data after it is sold is changing all of the time, as people come up with new ways to analyze and use data. Your data could be sold to an insurance company that could affect your premiums. It could also be sold to a real estate agency that may use your data to discriminate properties when showing homes. Politicians could use this information to sway the emotions of voters. When a user builds data, they are unsure of how it could be sold and used to later affect them in the future.

* It may be the case that to address these ethical questions effectively, we first need to eliminate some "conceptual muddles"—such what search engines are and how the law should treat them. In a recent paper ("Speech engines", SSRN, Apr. 7, 2013) James Grimmelmann considers three ways we might treat search engines:  
    
   1. as passive, neutral *conduits* of authors' speech,   
   2. as active, opinionated *editors*—speakers in their own right, or  
   3. as helpful, trustworthy *advisors*, much like reference librarians.  
    
  Each of these roles is associated with different ethical principles, laws, rules, and practices. Consider one of the legal or ethical questions you identified above, and compare how responses to this question might differ depending on whether we treat search engines as conduits, editors, or advisors. **(3 points)**

When a search engine automatically gives you documents to consume, the ethical and legal questions behind it will change dramatically. When a search engine acts as just as passive, neutral conduit of the documents online, there is a huge potential for a result to be misleading information. It puts the same amount of weight on documents that are scientifically backed and those that are not. For example, a document from a reputable website on how to deal with a cold could be returned just as likely as one that comes from a website that makes the claim that mercury is a good cure for the common cold. If these kinds of occurrences were common, the ethical question of whether or not the search engine would be responsible for pointing towards documents that provide potentially life-threatening advice.

When search engines act like opinionated editors, they will easily become political targets – tools that one party may use to affect elections or political opinion. If Google’s “search without searching” feature provided only one-sided opinions on an issue, it could have a significant impact on the public opinion of an issue. If the majority party in a country’s government agreed with Google’s stances on issues, they would most likely not interfere; however, if they disagreed, it would paint a target on the company for legislation to regulate search engines.

When search engines act as advisors, there brings an ethical issue of neutrality. Naturally, an advisor might want to point towards reputable, well known sources of information; however, this might lead to smaller, less known sources of information to be neglected. This is especially important in the commercial field of the Internet. A larger company may be very reputable and will be returned as a result to people, neglecting smaller, newer companies. Larger companies might use their vast amount of resources that smaller companies don’t have to influence the way that search engines return results to drive out competition.

## Bonus Question (2 points)

If you've made it this far and still have a little energy left, please tell me how this course has changed your understanding of information science by listing two things that you found most interesting, memorable or surprising about information science and why.

One of the most interesting topics was learning the fundamental way in which most information retrieval systems work. Most IR systems work collecting a corpus, indexing, retrieving information from that index, and ranking the search results. I found this interesting because of how normal it has become to use an IR system, to use anything from Google to Spotlight without truly understanding how the magic search box brings back useful results.

The second most interesting thing we went over in the course was the ethics behind information science. Ethics regarding information have had a colossal issue catching up to the science, and laws regarding them have been even further. Many people who use build information online (including lawmakers) don’t understand the ethics behind an information science issue and it requires a great deal of thought to come up with ethical solutions to problems. I am curious to see if there will ever be a standard for ethics that catches up with information science and usage.