Assignment 1:
Finding and Assessing Research Papers

Due Date: 09/30/2024 @ 11:59pm

The first step in research is getting yourself educated on the topic of interest. One of the most valuable sources for building background (assuming you already know what is in the related textbooks and you want to push your knowledge to the cutting edge of science and technology), is to read conference and journal papers on the topic of interest.

In this assignment, we learn how to find resources, and how to quickly evaluate them in terms of quality, relevance and perceived importance. Being equipped with this research skill saves you a lot of time, allows you to gather relevant reading material quickly, guides you to examine the quality and relevance of the reading material, and helps you spend your most valuable currency (your time) wisely.

How to find research paper?

There are several venues where you can find research papers related to Electrical and Computer Engineering. We cover some of these possibilities below:

1- IEEE Xplore Digital Library:

https://ieeexplore.ieee.org/Xplore/home.jsp

If you are on campus and connected to WiFi "GMU Secure", you can access IEEE Xplore without authentication. If you are off campus, you need to use VPN to authenticate yourself first as a GMU student, and then access the IEEE Xplore. Also, you can use GMU Library to access IEEE Explore through a browser (without VPN). When on the site, you may search for papers by typing keywords "e.g., low power Cache or Near-Threshold Computing", by entering the name of an author, etc. There are many filters that you could choose to limit the search results in many ways (e.g. year or relevance). You can also mix search criteria in the advanced search tab. For example, you may compose a search query by looking for an author name in the author list, and looking for a specific keyword in the text, and looking for another keyword in the paper's abstract.

GMU VPN access instruction could be found in the following page:
https://itservices.gmu.edu/services/view-service.cfm?customel_dataPageID_4609=6169

GMU Library allow you to access IEEE Explore through a browser (without VPN); go to the below website and type "IEEE Explore" in the search box.
https://infoguides.gmu.edu/dbs

2- **ACM Digital Library**

https://dl.acm.org/

Similar to IEEE Xplore, ACM Digital Library is a database of scientific publication with more focus on computing (hosting a range of publication of interest to CE and CS scholars). You can access ACM publication on campus for free. You can also login using VPN and authenticate yourself (as a GMU student) to access ACM publication off campus. Also, you can use GMU Library to access ACM publication through your browser (without VPN).

3- **Cornell University arXiv**

https://arxiv.org

arXiv is an archive and distribution server managed by Cornell University for research articles. Unlike IEEE Xplore and ACM digital libraries, the papers submitted to the arXiv are not peer-reviewed. Many researchers use arXiv to share their accepted conference and journal paper prior to the conference date or to report results obtained on a research project as a work in progress.

Assignment 1 (task 1.1):

Using IEEE Xplore list all publications of "Khaled N. Khasawneh" "Houman Homayoun", and "Avesta Sasan" that contain the keyword "Architecture" in the full-text. There is an option to create a citation for each of these papers from IEEE Xplore: Select all papers that fit the search criteria >> Select "Export" >> Select "Citation" >> Select "Plain Text" >> Select "Export". Copy and Paste the citations into your report. (do not copy the information after doi).

Assignment 1 (task 1.2):

Using ACM digital libraries list all publication that have the same search criteria as (task 1.1). Include in your report the total number of papers found using both IEEE Xplore and ACM digital libraries.

How to evaluate research papers?

In this section, we cover some of the basic evaluation tricks that you could use to narrow your reading material to more influential and likely more relevant work. Following are some of the approaches and metrics that you could use to filter or order the research papers that you have collected in the previous step.

1- **Evaluate the publication venue:**
(the conference or journal that the paper is published)

Conferences and Journals vary widely in terms of quality. A work that is published in a high-quality journal or conference is more likely of high quality (although many exceptions exist). In this part of the assignment, we learn how conferences and journals are ranked:

Journal Ranking:
Journals are evaluated using various metrics. Some of these metrics are discussed below: (source: https://latrobe.libguides.com/ld.php?content_id=36945524)

a- **H-index:** is computed by finding the maximum value of variable h, where h is the number of papers in a journal that has received h citations. The h-index combines the number of publication and number of citations into a single metric.

b- **H5-index:** h-index computed for articles published in the past 5 years.

c- **Journal Impact Factor (JIF)**: is obtained by dividing the total number citations to the publications in a journal over the total number of journal publications in the previous two years.

There are several other metrics (e.g. SJR. SNIP or Eigenfacor) that are also used for rating journal papers, description of which could be found in https://latrobe.libguides.com/ld.php?content_id=36945524. Note that these metrics are valid when comparing journals within the same discipline. You will find the listing and impact factor of many journals in the document below:

- Scimago Jounal and Country Rank (SJR)
  https://www.scimagojr.com
- GuideToResearch
  http://www.guide2research.com/journals/

Conferences Ranking:
When assessing the strength of a conference paper, pay attention to which conference the paper is published in. The higher quality conferences tend to have a higher bar (in terms of both quality and presentation of work) for a paper to be accepted. Note that each scientific field has an established reputation for various conferences related to that field, and usually the reputation is somewhat reflected in the acceptance rate of the conference. For example, ISCA is regarded as the highest ranked conference in Computer Architecture, and the acceptance rate of this conference is on average around 15%. However, note that publications in these conferences are usually highly cited. This makes publication in these conferences very desirable. The h-index and h5-index are frequently used to rate conferences. In addition, to these metrics, you can also find the following additional ranking systems useful:

a. **ERA:** Excellence in Research in Australia database: This ranking system is managed by Australian Computing Research and Education Association. The rankings range from A (=highest) to C (=lowest).
http://www.conferenceranks.com/data/era2010_conference_list.pdf

b. **Qualis:** This ranking system is published and maintained by the Brazilian ministry of education. Qualis uses the H-index as its evaluation metric. Qualis ranks the conferences from A1 (=highest), A2, B1, ..., to B5 (=Lowest). http://www.capes.gov.br/images/stories/download/avaliacao/Comunicado_004_2012_Ciencia_da_Computacao.pdf

c. **MSAR:** Microsoft Academic Field Ranking for conferences work similar to h-index and calculates the number of publications by an author and the distribution of citations to the publications. The field rating only calculates publications and citations within a specific field and shows the impact of the scholar or journal within that specific field.
https://academic.microsoft.com

A good place to start assessing the quality of the conference is the following link, where you get the ERA, Qualis and MSAR metrics in one stop:
http://www.conferenceranks.com/?searchall=IEEE#data

Assignment 1 (task 2):

Go to my webpage listed below, go to publications, and rank the all **conference (a.k.a symposium)** papers listed using each of the ranking systems above. Create a table of conference ranking using each ranking systems in your report. Table should be sorted from highest ranking to lowest ranking paper. (Note: a ranking for each conferences/symposium may not be found in all ranking systems) http://mason.gmu.edu/~kkhasawn

WARNING: The ranking of a conference should not be the basis for rejecting a paper for being a poor paper. You can think of papers accepted in higher ranking conferences of statistically carrying higher chances of being of higher quality or more relevant to the hot research topic at the time published. But note that many papers published in lower ranking conferences could be of high quality and may have a significant impact on the society.

2- Author influence and productivity

Another tool in your research toolbox is gathering information about the author(s). To learn more about the author(s), you can use a combination of search engines and metrics described below:

a- H-index: this metric combines the number of publication and citations received by published work of a scholar as a measure of productivity. The formal definition of h-index is as follows: If a scholar has an h-index of h, she/he has published h papers each of which has been cited by other scholars at least h times. For example, an h-index of 15 means that the scholar has at least 15

papers that are cited at least 15 times. The mechanism used for computing h-index is visualized in Fig. 1. A rough and quick way to compute the h-index is the following equation: h-index$= 0.54\sqrt{N}$, where N is the total number of citation for a scholar work. where N is the total number of citation for the scholar's publications. Note that sometimes the h-index is not a fair metric. For example, consider the following two cases. Researches A and B receive the following citations for the published work:

A → P1(1000) P2(198) P3(55) P4(50) **P5(5)** P6(2) P7(1) P8(0) P9(0)
B → P1(12) P2(11) P3(10) P4(8) P5(7) **P6(6)** P7(2)

The researcher A has several papers that are highly cited. A also has a larger pool of publication. But based on the h-index metric system A has an h-index of 5. The researcher B has a far smaller number of citations, and even a smaller number of publication, but has a higher h-index of 6.

The conclusion is not to rely on a single metric to evaluate the performance of a scholar.

(Reference: https://en.wikipedia.org/wiki/H-index)

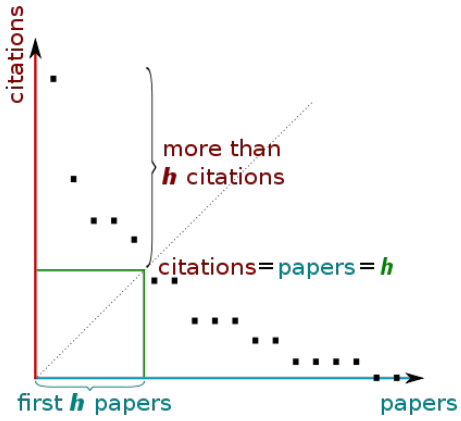

Figure 1 : computing h-index, source: https://en.wikipedia.org/wiki/H-index

b- I10-index: is the number of publications (including conference, journal, book chapters and books) of an author that have at least 10 citations. For the example above, author A has i10-index of 4 and author B has i10-index of 3.

NOTE: there are many other alternatives to these metrics. Some metrics try to qualify the quality of citations, while some other try to predict the future productivity of a scholar, and some other that intent to normalize the author's impact across different scientific fields. In my opinion, each of these metrics has its own value and flaws. It is the combination of many metrics and factors that should be considered collectively to measure a scholar's productivity.

There are several specialized search engines that could help you collect a lot of information on a scholar's publication and research activities. Some of this search engines include:

a- Google Scholar: https://scholar.google.com

Google scholar is in indexing toolbox designed by Google that indexes and automatically scans the scientific papers published in conferences and journals, and uses the reference section of each paper to count the number of times each previously indexed scientific paper is cited. The number of citation of an author is a good indication of how widely their work is received by the community. In addition, Google Scholar provides you with information on h-index and i10-index of the author for the past 5 years and also for the entire research career of an author. Note that the number of citations is not a metric to compare the quality of different researchers. Different research areas are cited at different paste (which roughly depends on the number of researches working on that research area). In addition, the number of citations tends to increase over time if an individual remains research active. Hence the comparison of the citation of a new faculty with only a few years of research with a full professor with many years of experience may not be a good comparison. However, you can sort the publication by year and observe how frequently the recent publications are cited. This probably provides you with a better judgment on how relevant and impactful the current research of the author is.

b- Dblp https://dblp.uni-trier.de/

Dblp is a good database that provides you with a quick overview of published work by an author. The coloring scheme used is extremely useful. The journal papers are marked red, Conference papers are marked blue, archive papers are marked grey, and books (and book chapters) are marked yellow.

c- SemanticScholar https://www.semanticscholar.org

Provides you with a quick overview of the author's research portfolio, displays a list of other authors that have influenced the author the most, and the researchers that are influenced the most by the work of this author. An interesting metric in semanticScholar is the Highly Influential Citations, which identifies citations where the cited publication has a significant impact on the citing publication, making it easier to understand how publications build upon and relate to each other. Influential citations are determined utilizing a machine-learning model analyzing a number of factors including the number of citations to a publication, and the surrounding context for each. (Note not all papers have Highly Influential Citations)

d- Microsoft Academic https://academic.microsoft.com

The Microsoft Academic provides you with a good snapshot of the author's broad research activities. Affiliation of coauthors and research topics investigated by the author in one shot.

3- Paper Citation

The number of citations of a paper is a good indicator of how influential that paper has been. Note that the number of citations should be considered along with the age of the paper. The number of citations usually remains low for about a year until the

paper is recognized and cited by other works that find the publication relevant or inspiring. Then the citation count grows for a while. When the topic is no longer a hot or research topic, the number of new citations drops over time. Fig. 2 captures the generic citation curve of a published paper. This figure is obtained from a study that looked at characteristics of papers published in Brazilian accounting journals (you can access the full paper at https://www.researchgate.net/publication/262800822/download) and the timelines may not apply to other science fields. Especially computer and electrical engineering fields in which the technology changes at a much faster paste. However, the general shape of the curve fits the trend of citations in various fields when averaged across many publications. Note that there are many exceptions, and the citation count of many papers follows a very different curve. For example, there are older visionary papers that keep a high citation count year after year, or there may be older solutions, that for lack of technology were not adopted at the time of publications, but after years or decades, they become relevant to the field.

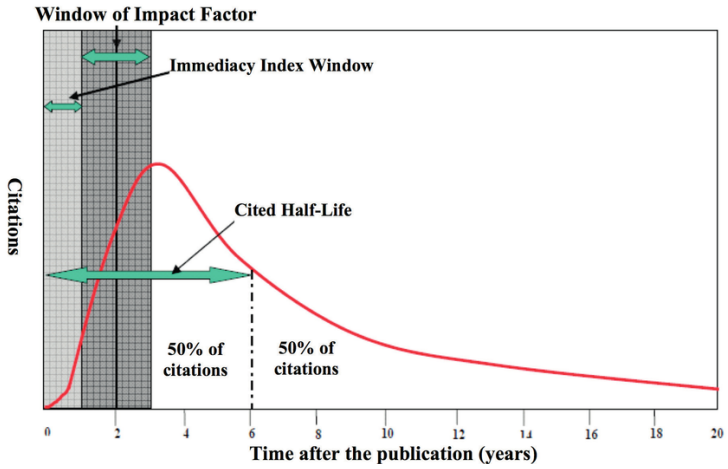

Figure 2: Generalized citation curve. Source: Amin & Mabe (2007).

To find the number of times a paper is cited, you may use search engines such as Google Scholar (**https://scholar.google.com)** or Semantic Scholar (**https://www.semanticscholar.org).** As discussed previously, the velocity of the citation (average number of citation per recent years) is also a good indication if the paper is still a highly influential paper. Combination of citation count and velocity could tell you a lot about a paper. For example, a high number of citation and low velocity would describe an older publication that was influential at some point in time in the past but is no longer an active research topic.

Assignment 1 (task 3):

Use Google Scholar to find the number of citations for each paper you ranked in Task 2 of Assignment 1. You can search for a part of the paper's title in google scholar to find the papers. Alternatively, you can search my name, sort the results you found based on year and find the papers.

4- Source of Research Funding

Usually, research papers that are supported by a funding agency acknowledge the support by adding an "Acknowledgement" section to the paper. If the research is supported by a reputable funding agency (such as National Science Foundation), you can assume that at least some content of the paper is vetted by peer reviewers during the proposal review process, indicating that the work is assumed to be of high quality by peer experts.

A Final Note:

The references cited in an influential and well-constructed paper are also good candidates for learning more about the topic and its background. When reading papers, try to memorize the name of authors. You probably forget them soon after you are done with that paper. But if the authors are active in that area, you will come around these names again and again. Follow the work of influential authors. The related work section on conference and journal papers also serve as a perfect mean for learning more about the community and related influential work. The more you know the community, the easier it becomes to quality your potential reading material.

Assignment 1 (task 4):

Scenario: You have signed up for Ph.D. and your advisor asks you to read papers related to "Cache Architecture". Then he asks you to find 10 highly influential papers on cache architectures and to identify 5 very active researchers in this area.

Your task:

1- Identify and list the highest influential papers. The papers should be cited in IEEE format. For each paper provide the number of citations as of date.

2- Identify and list the active researchers in this field. For each author provide the number of citation, H-index, and i10-index. Identity how many journals, conference papers, and book chapters (or books) are written by each author.

In this assignment, we learned some of the tools and evaluation metrics available to us to find scientific articles related to a topic of interest and how to quickly assess and rank the papers. This skill enables us to separate the more influential and possibly more informative publication as the first step in research. As you get to know the related research community more, it becomes easier and easier for you to identify and follow relevant and high-quality publications.

In the next assignment, we take the task of reading and reviewing research papers. We look at research papers form two different perspective. (1) How to learn from research papers,

and (2) How to review the research papers and assess their strength and weaknesses. In the rest of this class, you will be responsible to read assigned research papers, write summaries on these papers, and judge the strength and weaknesses of each paper. This exercise will provide you with an extremely valuable tool to build knowledge, regardless of the pathway you take to become an engineer or scientist.