Course unit: SCS 3108 - Data Structures and Algorithms

Module name/Topic: Ethical issues in Data Structures and Algorithms

Contributors: Christopher K. Chepken, Evans A.K Miriti and Andrew M. Kahonge

Duration 3 hrs

Reviewed by: Christopher Moturi

Module Description

This module is designed to introduce you to ethical issues in Computing and specifically data structures and algorithms. At the end of the module, you should be able to discuss ethics in data structures and algorithms. The module is divided into three main sections. The introduction section gives you the general understanding of how computing can be unethical. Section two talks about choosing algorithms and the general ethical concerns. The last part gives a few case studies which will provoke your thinking on the various ways in which sorting and searching algorithms may

Module learning outcomes:

By the end of this module, the learner should be able to:

1. Describe ethical concerns applicable to data structures and algorithms/
2. Explain why use of Computers and Computer algorithms may lead to unethical outcomes
3. Apply ethics in computer algorithms
4. Discuss the importance of ethics in algorithm design, implementation and use

Topic/Module content:

Introduction

Research on the ethical implications of algorithms, particularly in relation to fairness, accountability (does it capture every details for example in a priority queue environment where various groups of people are to be served based on priorities developed using several factors such as age, wellness, PWDs, arrival time, VIP etc), and transparency (Do we know how sorting and searching algorithms work, pregnant or lactating women,) This may lead to inconclusive evidence leading to unjustified actions.

It is important for any Computer Scientists or any algorithm developer or user to understand that using or infusing technology into an existing working social system may change the behaviors, understanding and other core values of the system. Equally important is the fact that the best solution to a problem may not require any technology and hence forcing technology may become unethical in the first place. For example:

1. Sorting algorithms may introduce some sort of discrimination i.e. the concept of sorting is to separate elements (entities which could be people). If a particular sorting principle is applied, it may end up discriminating against others.
2. Searching algorithms in some cases may lead to isolation.

Even though computer algorithms are designed to be applied in a fair manner to real world problems, there is need to understand the general public's understanding of fairness. In addition, even though data algorithms may be fair, the data used may be biased and hence as an algorithm designer and user, it's your responsibility to interrogate the data in use. The design and coding of the algorithm should also be fair enough to a point that it can factor in extraordinary situations. For example, a betting algorithm which chooses a particular number as a winner (say the 17^{th} person to send an SMS will be the winner). What happens if two messages come in at the same time? Does the algorithm allow for two winners to be announced?

Choosing Algorithms

What informs an algorithm implementer or developer on the choice of an algorithm will vary from person to person and situational circumstances. Among the important factors are one's knowledge

on the algorithms, the context of use, the sensitivity and accuracy of the algorithm output among many options. As a person in charge of choosing the algorithms and data for a particular problem, its ethical that you have the required skills and knowledge to execute your mandate.

General ethical concerns

. Algorithm applicability: Will one sorting, searching or categorization algorithm be universal? For example categories of people: all genders, races, PWDs, marginalized communities among many other categories.

a. Explain ability of algorithms to different groups

b. Fairness and Favoritism: Refocusing search and sorting algorithms to suit a particular context. For example, you may want to bias it to favor PWDs. Think of an algorithm used to allocate contracts to contractors. In a case where you have to favor youth and women (as required by the law of Kenya), one may need to:

.Explain to the others how the biased algorithm works

i. The algorithm may become too biased or may not serve the purpose correctly

ii. Ensuring fairness even when an algorithm is used for different context areas.

a. Inconclusive evidence: Under some circumstances, results from any Computer algorithm may produce none binary results. At that point, ethical issues may arise

b. Inscrutable evidence (lack of transparency): In some cases, it is not clear how the algorithm arrives at the results. For any good algorithm and data, it must be clear to the users and developers how a certain result was arrived at.

c. Misguided evidence: Due to various reasons, such as noisy data, some algorithms may produce incorrect results.

d. Unfair outcomes- leading to bias and discrimination

e. Transformative effects – leading to challenges or autonomy and informational privacy. Algorithmic decisions do not have a second chance decision. Autonomy of users is also reduced by the transformative algorithms

f. Traceability: Using any algorithm and data must allow for traceability.

Supplementary reading materials

In this section, we have provided a quick recap on Sorting and searching algorithms. This reading assignment is meant to remind you of the basics of sorting and sorting algorithms but makes the assumption that you have learnt the same in your previous lectures.

Sorting

Sorting is the arrangement or organization of data elements in a desired order. It is required to optimize searching. In data structures and algorithms, it is important to learn sorting algorithms for the following reasons:

. Using various sorting algorithms, we learn various problem solving approaches

a. sorting is used to solve several coding problems

b. It is easier to learn analysis of algorithms using sorting algorithms

Examples of the problems which may require sorting are

1. Looking for a pair of items which are a subset of a bigger set of items
2. Getting the most frequent appearance of an element in data e.g. for a character in an array
3. Check if two sets of data are equal; Think of analyzing health related data and you are required to find out if a certain region has the same characteristic as another.
4. Check if some records are missing: For example, in recording student exams, the ordering of student registration numbers in ascending or descending order could help identify a missing student.

Searching

Searching in data structures is the process of finding desired information from a set of elements stored in the form of elements in a computer's memory. It is the process of identifying a desired element with certain characteristics within a collection of elements. These sets of elements are organized in different data structures such as arrays, trees, graphs, and linked lists.

A search algorithm finds or retrieves elements in any form of the stored data structure. Categorized into two based on how searching takes place.

1. **Sequential Search:** An array or list of elements is traversed sequentially, examining each component of the set. Ex. Linear Search
2. **Interval Search:** Algorithms designed explicitly for searching sorted data structures are included in interval search. The efficiency of these algorithms is much better than linear search algorithms. For example, binary search, logarithmic search.

There are different types of search algorithms, some of which are more efficient than others. Others work better with some certain problems, data sets and different contexts. To be able to identify a good search algorithm, search algorithms Algorithm analysis is carried out to find the amount of time an algorithm will generally take to find the element(s) being searched. This largely depends on the kind of data structure used to store data and whether the data is sorted or not. Examples of search algorithms are linear and binary search algorithms.

Real world applications of sorting and searching algorithms

e-commerce	Can be used by customers to find and compare products based on a criterion such as price, rating, category, or availability. Quick sort and binary search algorithms are used.
social media,	Can help users find and interact with content and people based on their preferences, interests, or connections. Merge sort and hash table search algorithms are ideal.

education	Users can organize and access information and resources based on their needs, goals, or levels. Insertion sort and depth-first search algorithms can be used for this purpose.
health care	Doctors and patients can use sorting and searching algorithms to diagnose and treat diseases and conditions based on their symptoms, tests, or treatments. Heap sort and breadth-first search algorithms can be used for this purpose.

The use of sorting and sorting algorithms can be combined in a way that forms a particular algorithm, largely as AI algorithms such as rule based, fuzzy logic and machine learning algorithms.

Searching and sorting algorithms challenges

a. Complexity leading to lack of understanding and transparency
b. trade-offs between time and space complexity, stability and instability, and simplicity and sophistication
c. Can have some limitations and drawbacks in terms of data quality, security, and privacy, and thus require proper handling and protection of data and information.

Cases

Ethical issues cases in sorting and searching algorithms

Case one: Class discussion

a. A system was developed to select an individual for a job. With some modification the same algorithm can be used to identify crime suspects. The system returns the first person that is found.
    i. What happens to the others with similar characteristics?
    ii. Will it get the correct selection?
    iii. Is there a possibility that we select a non-criminal and brand them as a criminal?
    iv. Discuss any other ethical issues which may arise.

Case two: The same characteristics

(A case for ranking students with similar marks)

What is the effect of having sorting which requires ordering in some way and priority numbers given and yet there is a possibility of having the value being used to sort shared by many elements?

a) Will the algorithm be fair?
b) Will the algorithm be consistent?

Case three: Bias and discrimination

Consider a search algorithm used in machine learning to identify specific characteristics which may not be available in some groups of people or communities. This may bring bias hence discrimination.

a) Cases of police criminal prediction algorithm: in such cases, investigation agencies may employ algorithms which may inadvertently categorize someone in a wrong group. This will raise ethical issues:
b) where predictions made by algorithms provide no guarantee that they are right.
c) The investigators acting on incorrect predictions may be forced to form unjustified or biased investigations guided by the algorithm outcome.

Case four: Gendered language

Consider, for example, how the outputs of translation and search engine algorithms are largely perceived as objective, yet frequently encode language in gendered ways (Larson 2017; Prates et al. 2019).

Assignments /Quizzes:

1. Which one of the following is a key characteristic of sorting algorithms?
    A. It is used to arrange data elements in ascending or descending order {Answer}
    B. It is used to analyze algorithms in algorithm analysis
    C. A sorting algorithm can be used as a searching algorithm.
    D. Sorting algorithms can be used as a problem solving strategy
2. Which of the following best describes the relationship between the quality of data used in searching algorithms and fairness?
    A. Using biased data may introduce unfairness in a search algorithm {answer}
    B. Clean data will always produce unbiased search results
    C. Unclean data will always produce biased search results
    D. A fair algorithm is cannot be affected by the data used.
3. Is the statement "searching algorithms in some cases may result into isolation" true or false
    A. False
    B. True {Answer}
4. Which one of the following statements is FALSE about forcing technology on a system?
    A. It may force people within the system to do unethical things
    B. It may benefit the system in a way they didn't anticipate
    C. It will always lead to unethical system {answer}
    D. The technology may not have any negative or positive impact
5. Which one of the following statements best describes algorithm applicability as an ethical issue in algorithms?
    A. Due to various reasons, such as noisy data, some algorithms may produce incorrect results.
    B. Search or sorting algorithms may lead to bias and discrimination
    C. Use of any algorithm and data must allow for traceability

D. There is a need to interrogate any sorting, searching or categorization algorithm for universal application. {answer}

6. With the understanding of sorting algorithms and their principles, identify any ethical issue, principles and values known to you as a learner.

References

Ethics of Algorithms (cihr.eu) Center for Innovation in Human Resource. (2015). Ethics of algorithms. https://cihr.eu/eoa2015web/

Real-Life Application of Sorting and Sorting Algorithms. (n.d.). https://www.linkedin.com/advice/3/how-do-you-apply-sorting-searching-algorithms

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). The ethics of algorithms: key problems and solutions. *Ethics and Information Technology*, 23(3), 215-223. https://doi.org/10.1007/s00146-021-01154-8