Thomas Swed
Research in Computer Science – CPSC-59700
3/31/2017

Algorithm analysis is a way to measure the complexity and speed required to execute a program. It compares differences to determine the most optimal idea or algorithm. Big-O notation is often used among computer scientists to indicate the "worst case" or "upper bound" of an algorithm in terms of its complexity. An algorithm that is found to have the least amount of complexity is said to be optimal, meaning all other algorithms have a worse or equal amount of complexity. By performing algorithm analysis, we gain a better understanding of a given algorithm and how it relates to other algorithms. Algorithm correctness is related to the input-output of a program mathematically speaking. It is said to be correct if an algorithms output can be expected or "proven" based on a given input. For all possibly input, the program must also terminate. This can be done using formal proof technics. Distributed Algorithms are algorithms designed to run on separate processors/nodes while communicating. Because these algorithms are distributed, distinct parts of the algorithm are run simultaneously on separate processors that have limited information about the rest of the program. While the nodes communicate with each other, each node only sees in part rather than the whole. The design of a distributed algorithm deals with how nodes communicate with each other effectively and efficiently to execute a singular program. Because of this, distributed algorithms are especially difficult to design but have several advantages when compared to a centralized system.

Thomas Swed
Research in Computer Science – CPSC-59700
3/31/2017

I work for a book publishing company in Wheaton, IL. There are several departments within the company that utilize a number of computer science topics. Among them are security, artificial intelligence and perhaps the most obvious, databases.

Every company has some type of underlying data that it is required to account for, secure, and query. My company uses a MySQL relational database which has its advantages as well as setbacks. Some problems with relational databases include their inability to handle unstructured data efficiently, such as searching for a term or phrase in a product description. They provide little assistance when dealing with unexpected formats and don't scale well to large applications. Other challenges arise when the database is replicated – the process of copying data from one server known as the master to one or more different servers known as the slaves. Replication increases security by creating backups, improves performance as the load is now distributed among multiple servers and provides flexibility in executing analytics as they can be performed on the slaves rather than the master server. To achieve these benefits an efficient way to implement this replication is needed, this presents issues in determining the needed bandwidth and resources. Relational databases were designed to run on a single server and as such cannot compare with the speed, scalability and reliability of distributed databases. They are not predictable or reliable in terms of performance as any given query could consume a large amount of resources or a small amount depending on the complexity of the query, indexes and how the database is maintained. The duration or performance of a query is also dependent on any other queries running at the same time it is being executed.

While there are advantages to a distributed database, there is always the other side of the coin to consider. Many problems can be solved in turn but further research is needed to determine the best type of database for an organization's needs. Distributed databases require complex software and can be difficult to implement. Deadlocks are difficult to account for as compared to a centralized system and data integrity becomes complicated, requiring much network resources. Further research is required to compare and analyze differences in design and discern the specific goals needed within the company. Once implemented, research in overcoming previously mentioned challenges is also needed.

Another important factor in terms of data is security. Important and sensitive data such as credit card information should have the highest level of security protocols applied. If a third-party payment provider is not used, PCI Guidelines must be followed (legally speaking). This provides its own set of challenges as PCI Guidelines are strict and difficult to adhere to for the average company. Small businesses are often faced with a smaller budget dedicated to things like IT and security. Educating staff on simple technical precautions may also seem like a daunting but necessary task. By researching emerging new technologies, a cost effective and efficient way to set up a secure network, train staff and adhere to guidelines can be achieved.

Artificial intelligence is also being explored in the Marketing department of my company to aid in the analysis of social media to determine specific wording to use in product descriptions and blog posts. The desired result of this tactic is to get more search hits from a broader audience by utilizing commonly searched for terms. AI is also being explored/researched to be used in our Editorial department to analyze the content of best-selling books. The patterns and literary styles found in best sellers can then be applied in a general sense to books that have not been published yet. This presents coding challenges as well as the time and resources needed to implement the software. Through research perhaps an efficient algorithm can be developed that would reduce the time needed to accomplish AI programming while producing desired results.