Imani Palmer

7/19/2013

Donald Ervin “Don” Knuth

Donald Ervin “Don” Knuth was born on January 10th, 1938, in Milwaukee, Wisconsin to Ervin and Lousie Knuth. In secondary school, Don entered a competition set up by Ziegler, a confectionary manufacturer. The goal of the contest was aimed to see how many words could be made with the letters of “Ziegler’s Giant Bar.” Don spent two weeks and found 4500 words. The judges of the competition had only found 2500, Knuth was the clear winner. In high school, Knuth had a growing passion towards music. Don played the saxophone and the tuba in the school band, as well as composed music. Don graduated from high school with the highest grade point average ever achieved at his high school. After receiving a scholarship to attend Case Institute of Technology in Cleveland, Ohio, he decided to major in physics.

During the summer between his freshman and sophomore years, Knuth worked in the statistics lab drawing graphs, punching key cards, and using a card sorter. At this time, he spotted the newly installed IBM 650 computer. Don consulted “... the manual [he] got from IBM…” from the programming examples in the book, Don “knew [he] could do… better than that...” In his sophomore year, he made the official decision to switch his major from physics to mathematics. Knuth used his growing expertise at writing computer programs to analyze the performance of the college basketball team. It led to some publicity and IBM used a photograph of him in their advertising. On June 1960, Don Knuth graduated with his Bachelors of Science and was also rewarded with a Masters. Knuth was also awarded two fellowships, a Woodrow Wilson Fellowship and a National Foundation Fellowship. Knuth decided to attend the California Institute of Technology in order to purse his PhD in mathematics. While in graduate school, Knuth consulted and wrote compilers. In January 1962, Addison-Wesley asked Knuth to write a book about compilers. Don sketched twelve chapters and signed a contract. This was the beginning of Don’s long career as an author.

During his time as a PhD student, Don read many technical articles, and noticed the spotty and sometimes unreliable nature of the literature in the field of computer science. He saw the need for someone to write a book which organized and reliably presented what was known this task. Knuth was a good writer and had an instinct for trying to organize things, so he decided to tackle. He used a quantitative rather than qualitative approach, and emphasized the creation of programs that are beautiful. The book reaching 3,000 hand-written pages, Addison-Wesley decided that the book should be reorganized into 7 volumes, with a chapter or two per volume. The first four volumes are based on concepts and information structures, random numbers and arithmetic, sorting and searching, and combinatorial algorithms. Volumes 5-7 are based on more compiler specific chapters such as lexical scanning and parsing, context free languages, and compiler techniques. Volume 1 of The Art of Computer Programming was published in 1968. By 1973, Knuth had published volumes 1-3 of The Art of Computer Programming (“TAOCP”). TAOCP emphasized a mathematical approach for comparing algorithms to find out how good a method is. Arguably, the books established analysis of algorithms as a computer science topic in its own right. Knuth has stated that developing analysis of algorithms as an academic subject is his proudest achievement. Following his success in writing, many other accomplishments followed.

The first three volumes of TAOCP had great impact on the field and encouraged many people to build on his work. In 1974, Knuth was awarded the Association for Computing Machinery's Alan M. Turing Award. B.A. Galler commented, “The 1974 A.M. Turing Award was presented to Professor Donald E. Knuth of Stanford University for a number of major contributions to analysis of algorithms and the design of programming languages, and in particular for his most significant contributions to the ‘The Art of Computer Programming’ through his series of well-known books.” The collections of technique, algorithms, and relevant theory in these books have served as a focal point for developing curricula and as an organizing influence on computer science.

Knuth has made many contributions to mathematics and computing. One particular contribution is the Knuth-Bendix algorithm, an algorithm for computing with algebraic structures, particularly with groups and semigroups. This important contribution, published jointly with his student Peter Bendix in 1970, attempts to solve the word problem in algebraic systems by deriving consequences of given relations to give a complete set. Another contribution is Knuth's invention of TeX, a language for typesetting mathematical and scientific articles. Starting in 1976, Knuth took ten years off his other projects to work on the development of TeX and METAFONT, a computer software system for alphabet design. Consequently in 1977, Knuth began developing a new typesetting system to enable high quality computerized typesetting, in particular for TAOCP. This system was announced in his 1978 American Mathematical Society Gibbs Lecture entitled *Mathematical Typography*. Knuth had two goals for his system. The first was achieving the finest quality printed documents. The second was creating a system that would be archival in the sense that it was independent of changes in printing technology to the maximum extent possible. Knuth's system, developed with help from Stanford students and colleagues, had three primary components: the TeX typesetting engine, the METAFONT font design system, and the Computer Modern set of type fonts.  Combined, these revolutionized digital typesetting. Knuth made his code publicly available, and it has been widely adapted by commercial typesetting systems. Knuth put hooks in his TeX engine so that others could make additions, with the condition that any resulting system be give a different name. That produced a vibrant, worldwide community of users and developers for TeX and related systems like LaTeX, ConTeXt, LuaTeX. Knuth's TeX was an early success story for the free and open-source software movement. Knuth thought his typesetting work would take a year or two, but it was not until 1990 that he announced that he would make no further changes to his systems except to correct serious bugs

The KMP Algorithm is a linear time string matching algorithm. The KMP algorithm was conceived by Donald Knuth and Vaughan Pratt and independently by James H. Morris in 1977. KMP keeps the information that naïve approach gathered during the scan of the text. By avoiding this waste of information, it achieves a running time of O(m+n). Consider an example so that the algorithm can be clearly understood.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| String | B | **A** | | **B** | | C | | A | | B | | A | | B | A |
| Pattern | | | **A** | | **B** | | A | | B | | A | |

Execute the KMP algorithm to find whether the pattern occurs in the string. Compare the first index of the pattern, with the first index of the string. Since these characters do not match, it will be shifted one position to the right.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| String | B | **A** | | **B** | | C | | A | | B | | A | | B | A |
| Pattern | | | **A** | | **B** | | A | | B | | A | |

Now, compare the first index in the pattern with the second index in the string. Next, compare the second index in the pattern with the third index in the string, they do not match. Backtrack on the pattern, and compare the first index of the pattern with the fourth index of the string. They do not match.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| String | B | A | | B | | **C** | | A | | B | | A | | B | A |
| Pattern | | | **A** | | B | | A | | B | | A | |

Once again, backtrack on the pattern and a comparison of the first index in the pattern with the fifth index in the string. Proceed through the pattern, and match all of the remaining letters in the pattern. A match has been found.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| String | B | A | | B | | C | | **A** | | **B** | | **A** | | **B** | **A** |
| Pattern | | | **A** | | **B** | | **A** | | **B** | | **A** | |

The total runtime of this algorithm is O(n+m). The advantages of KMP are the linear runtime and the fact that the algorithm never need to move backwards in the input text. The disadvantages of KMP is that it doesn’t work so well as the size of the alphabet increases, which increase the chance of a mismatch occurring.

Knuth believes his greatest achievements were working “…on the theory that underlies algebraic languages.” He feels, “…the biggest thing that [he] developed was the mathematical approach to compare algorithms in order to find out how good a method was.” At Stanford he supervised more than 30 PhD students. He created important courses such as *Analysis of Algorithms*, *Concrete Mathematics*, and the legendary *Programming and Problem Solving Seminar*. In addition to his own teaching and research, Knuth has served the computing community on professional society committees, as an invited lecturer on many occasions, and on the editorial boards of more than 30 technical journals. He is the holder or co-holder of 5 patents. Knuth considers TAOCP his masterwork, and in 1993 he retired early to spend more time writing additional volumes. He had produced revised editions of volumes 1-3 in 1978-1979. He designed a new hypothetical computer to replace the MIX computer of volumes 1-3 for the analysis of algorithms; this new computer was described in Knuth's 1999 book devoted to the topic. Knuth began releasing volume 4A of TAOCP, on combinatorial algorithms, in fascicles ranging from 128 to 250 pages. In early 2011 the 921-page volume 4A was published, but the later parts of chapter 7 were reserved for future volumes. Donald Knuth is one of the preeminent computer scientists of our time. He has made major contribution to many areas, in effect pursuing multiple simultaneous and serial careers, any one of which would be a proud lifetime achievement for other people. He credits much of the success of his work to combining theory with practice. Knuth is the rare theoretician who writes many lines of code every day. Programming is an art he practices often.

References

"The Art Of Computer Programming." Interview by Donald E. Knuth. *The Chartered Institute for IT*. BCS, June 2011. 19 July 2013. <http://www.bcs.org/content/conWebDoc/40462>.

"Donald ("Don") Ervin Knuth." *Donald E. Knuth*. 19 July 2013.

"Donald Ervin Knuth." *Knuth Biography*. 19 July 2013.

Greene, Daniel H., and Donald Ervin Knuth. *Mathematics for the Analysis of Algorithms*. Boston: Birkhäuser, 1990.

Knuth, Donald E., James H. Jr. Morris, and Vaughan R. Pratt. *Fast Pattern Matching in Strings.* Stanford: Computer Science Department, Stanford University, 1974.

Knuth, Donald Ervin. *Literate Programming*. Stanford, CA: Center for the Study of Language and Information, 2001.

Mandumula, Kranthi K. "Knuth-Morris-Pratt Algorithm." 18 Dec. 2011.