



Cyril W. Cleverdon, Jack Mills, E. Michael Keen : Factors Determining the Performance of Indexing Systems ; Volume 1 : Design

The College of Aeronautics, Cranfield, England (1966)
<http://sigir.org/resources/museum/>

実験に基づき論文を書くのは当たり前?

「有名論文」ナナメ読みという企画だが、かなりマニアックな論文（というか Volume 1 だけで 120 ページもある報告書）を選んでみた。半世紀以上前に書かれたこんなに長い英文をいまさら読む日本人はほとんどいないと思うが、この中で報告されている Cranfield II と呼ばれる実験は、Web 検索エンジンなどの技術の根幹をなす研究分野「情報検索」におけるマイルストーン的成果である。

今日、データを用いて実験を行い、実験結果に基づき考察を行い、研究発表を行うのは当たり前である。しかし、少なくとも情報検索分野では、Cyril Cleverdon らによる Cranfield 実験以前にはこのような実証主義的アプローチが確立されておらず、憶測に基づく哲学的議論が主流だったらしい¹⁾。実際、Cleverdon は本報告書 p.10 において、実験をせずに手法の優劣を「300 ページにもわたって」論じた研究者のことを名指しで痛烈に批判している。

Cranfield II とは

以下、Cranfield II の中身について簡単に紹介するが、詳細については拙著²⁾ 第 1 章「情報検索評価の歴史」をご覧ください。

Cranfield II はその名の通り Cranfield I という実験の続編で、英国 Cranfield にあったカレッジ (現 Cranfield 大学) で実施された。その研究課題は、「図

書館にある各文献にどのような体系で索引をつければユーザが効率的に文献を探し当てられるか」であった。

索引とは何か？ たとえば Web 検索エンジンは、高速な検索を実現するために各 Web ページ内のテキストに含まれる単語をあらかじめすべて抽出し、「各単語がどの Web ページに含まれるか」を示す表 (転置ファイル) を自動生成する。これは索引の一種で、索引中に列挙されている語を索引語という。転置ファイルの索引語は Web ページに出てきたものであれば何でもよく、自由語と呼ばれる。これに対し、古典的なアプローチでは、検索のための用語やコードの体系をあらかじめ決めておき、各文献に対し必ずこの中から選んだ索引語を付与する。このような索引語を統制語という。人間が主観的判断により統制語を文献に付与するのが 1960 年代における常識であった。

Cranfield II は、人手で索引づけを行う際、統制語や概念体系を用いるよりも自由語を用いる方が有効であることを示した。これは、最も単純な方法が最も有効であるという当時としては驚くべき結果であり、情報検索研究が実証主義の方向に舵を切るきっかけとなった。Cranfield II はさらに、検索結果の質を定量評価するために、各検索課題 (どのような文献を探したいか) に適合する文献、すなわち正解データをあらかじめ人手で収集するアプローチを確立した³⁾。このプロセスを適合性判定という。今日では、検索対象の文書集合と、検索課題の集合

と、各検索課題の適合性判定結果からなるデータセットをテストコレクションと呼ぶ。

余談だが、筆者が「情報アクセス評価基盤」という大学院科目で Cranfield I, II を紹介する際、学生に「1960年代に行われたこの検索実験には、どんなコンピュータが使われたか？」とクイズを出すことがある。まじめな学生はまじめに考えてくれるのだが、これはひっかけ問題である。Cranfield I, II の全工程は人手で実施され、コンピュータの出番はなかったからである。

Cranfield II のインパクト

Cranfield II テストコレクションは、米 Cornell 大の Gerard Salton 教授や英 Cambridge 大学の Karen Sparck Jones 教授らの研究にも使われた。その後、研究者間でテストコレクションを共有し、同じ土俵で検索アルゴリズムを比較評価する研究手法が浸透していった。1992年には、米国において大規模テストコレクションの構築と共有を伴う検索評価会議 TREC (Text Retrieval Conference) ^{☆1} が立ち上げられ、今日でも続いている。日本では、1996年および1998年に初めて日本語検索テストコレクションが公開されたのち ⁴⁾、1999年に第1回 NTCIR (当時 NACSIS Test Collection for Information Retrieval, 現在 NII Testbeds and Community for Information access Research) ^{☆2} という TREC と肩を並べる国際会議が開催されている。NTCIR は2019年の第14回会議で20周年を迎えるため、これを記念した本が2020年に出版される。

Cranfield II のインパクトは学界にとどまらない。Web 検索エンジンを運営する企業では、今日でも人手による適合性判定結果をクリックデータなどの暗示的なシグナルと併用し、検索エンジンの評価と

調整に活用している。なお、このためには二値適合性 (文書が情報要求に適合するか否か) に基づく適合性判定では不十分であり、多値適合性 (文書が情報要求にどの程度適合するか) に基づく適合性判定が用いられる。5段階の多値適合性を採用していた Cranfield II はこの点においても先駆的であった。

Cleverdon は1997年に亡くなったが、1991年に ACM SIGIR より情報検索分野で最も栄誉のある Salton 賞を受賞している ^{☆3}。その受賞記念論文 ⁵⁾ には、Cleverdon が Cranfield I, II の結果を公表した当時、ほかの研究者たちからいかに批判されたかが綴られている。本稿で紹介した報告書は1966年のものであるから、当時「炎上」したことを四半世紀にわたり根に持っていた計算になる。情報検索研究にはこれくらいの執念深さが必要なかもしれない。

参考文献

- 1) Cleverdon, C. W. : Report on the First Stage of an Investigation onto the Comparative Efficiency of Indexing Systems, The College of Aeronautics, Cranfield, England (1960).
- 2) 酒井哲也: 情報アクセス評価方法論, コロナ社 (2015).
- 3) Robertson, S. : On the History of Evaluation in IR, Journal of Information Science, 34(4), pp.439-456 (2008).
- 4) 酒井哲也 ほか: 情報検索システム評価のためのテストコレクション, サイエンス社, Computer Today, 9(87), pp.31-35 (1998).
- 5) Cleverdon, C. W. : The Significance of the Cranfield Tests on Index Languages, Proceedings of ACM SIGIR '91, pp.3-12 (1991).

(2019年2月20日受付)

^{☆3} <http://sigir.org/awards/gerard-salton-awards/>

^{☆1} <http://trec.nist.gov/>

^{☆2} <http://research.nii.ac.jp/ntcir/index-ja.html>

.....
酒井 哲也 (正会員) tetsuyasakai@acm.org

1993年早稲田大学理工学研究科修士課程修了。博士(工学)早稲田大学。Microsoft Research Asia など通算約20年の企業経験を経て、現在早稲田大学教授、情報理工学科主任、情報理工・情報通信専攻主任。本会論文賞(2回)、山下賞、FIT論文賞・船井ベストペーパー賞など受賞。ACM Distinguished Member.