Jointly published by Akadémiai Kiadó, Budapest
and Kluwer Academic Publishers, Dordrecht

Scientometrics,
Vol. 51, No. 3 (2001) 607–637

Author-centered bibliometrics through CAMEOs: Characterizations automatically made and edited online

HOWARD D. WHITE

College of Information Science and Technology, Drexel University, Philadelphia, PA (USA)

This article describes ways of automatically generating 15 kinds of personal profiles of authors from bibliographic data on their publications in databases. Nicknamed CAMEOs, the profiles can be used for retrieval of documents by human searchers or computerized agents. They can also be used for mapping an author's subject matter (in terms of descriptors, identifiers, and natural language) and studying his or her publishing career. Finally, they can be used to map the intellectual and social networks evident in citations to and from authors and in co-authorships.

Introduction

The professional interests of publishing scientists – indeed, of writers in general – can be more or less accurately modeled through profiles that are automatically generated from terms associated with their writings in bibliographic databases. Existing technology makes it easy for anyone to create such profiles. Given appropriate software, the searcher forms the set of writings by an author and then asks for terms (noun phrases of a particular kind) that co-occur with the input name in bibliographic records, ranked high to low by their frequencies of co-occurrence. This simple process creates profiles ready for editing and use. Such profiling can be carried out with the software of several well-known providers of databases – for example, Dialog, Questel-Orbit, DIMDI, and the European Information Network Services. The more prolific the author, the richer the profile.

To give an immediate example, Table 1 is a profile of Steven Pinker, the well-known MIT cognitive scientist. It was automatically created in August 2000 from the subject headings for Pinker's books in the LC MARC-Books database, which covers monographic publications catalogued by the Library of Congress since 1968. (Pinker's 13 bibliographic records include some multiple entries for bestsellers like *How the Mind Works* and *The Language Instinct.*) Readers who have followed Pinker through the years will find that the list prioritizes his favorite topics quite plausibly. Moreover, it

0138–9130/2001/US $ 15.00

contains 19 exact phrases whereby anyone can pursue interests like his in the many databases that use LC subject headings to represent content. In creating the profile, one need not know the phrases or look them up; one simply exchanges Pinker's name for them during a Dialog search, like someone cashing in a chip in a casino.

The rank-ordering of terms yields core-and-scatter distributions of the sort long studied in bibliometrics. The distributions are often dynamic: as new writings from an author are captured in bibliographic records, new co-occurring terms may be tallied for the first time (as "onesies"), and the co-occurrence counts of terms previously included may increase. These counts are term weights. Since weighted terms are a lingua franca between bibliometrics and document retrieval, the profiles thus made have implications for both. But note that profiles of this nature personalize bibliometrics in a way rarely seen before, in that they center the core-and-scatter distributions not on a discipline, specialty, topic, or journal, but on a single author, construed as either an individual or an oeuvre.

Table 1
A subject-heading profile of Steven Pinker in LC MARC – Books

Rank No.	Items	Term
1	5	LANGUAGE AND LANGUAGES
2	4	BIOLINGUISTICS
3	4	LANGUAGE ACQUISITION
4	2	GRAMMAR, COMPARATIVE AND GENERAL
5	2	SEMANTICS
6	1	CHILD PSYCHOLOGY
7	1	COGNITION
8	1	COGNITIVE NEUROSCIENCE
9	1	CONNECTIONISM
10	1	HUMAN EVOLUTION
11	1	HUMAN INFORMATION PROCESSING
12	1	LEARNING ABILITY
13	1	NATURAL SELECTION
14	1	NEUROPSYCHOLOGY
15	1	PSYCHOLINGUISTICS
16	1	PSYCHOLOGY
17	1	SYMBOLISM (PSYCHOLOGY)
18	1	VERB
19	1	VISUAL PERCEPTION

We currently have the capability for a full-fledged bibliometrics of persons that is cognate with ego-centered analysis in social networks research (*White*, 2000). It is also cognate with the creation of personal agents that can browse indexing spaces on an author's behalf. For reasons not merely egocentric, authors tend to be interested in such things. Non-authors might find them interesting as well, especially when they involve established figures, from leaders in various specialties to world immortals.

Anyone who has published and been indexed can be rapidly profiled in way that leads, at minimum, to

- immediate or long-term search strategies for the retrieval of documents;
- maps of the author's subject matter in various vocabularies;
- maps of the author's relationship with other authors who are socially and/or intellectually connected ("ego-alter relationships" in the words of social networks analysts); and
- studies of individual publishing careers over time.

While all this may be recognizable enough to sophisticates, automatic personal profiling with actual content has been little discussed in information science in recent years. The relevant techniques are apparently little known even among professional online searchers, such as the information specialists and reference librarians who serve scientists and other learned clienteles (cf. *Wormell*, 1998a); and the clever people who are reinventing information science for the Web are often unaware of existing bibliographic systems and resources. I propose here to give examples of the varieties of this kind of profiling and to advance a terminology for its concepts, thus providing a base for further work in the area. Several of my examples involve Belver C. Griffith as focal author, in keeping with his place of honor in this issue of *Scientometrics* (see also *White* and *McCain*, 2000).

The software employed is Dialog's, the company that provides the fullest set of databases for author-centered profiling. However, my reliance on Dialog and its RANK and TARGET commands is merely a convenience; systems that rank-order terms by frequency have multiple implementations and will doubtless see increased use on the Web. Introductions to Dialog's RANK and the equivalent GET command on Orbit-Questel appear in *Snow* (1993) and *Bjorner* (1990). Papers that have explored capabilities for online bibliometrics include *Christensen* and *Ingwersen* (1996), *Hudnut* (1993), *Ingwersen* and *Christensen* (1997), *Jakobiak* (1985), *Manni* and *Serrazanetti* (1987), *Moed* (1991), *Persson* (1986, 1988), *White* (1989, 1990, 1996), and *Wormell* (1998a, 1998b).

Naming the profiles

Since 1993 I have informally used the acronym CAMEO, standing for Characterizations Automatically Made and Edited Online, to designate the general class of bibliometric profiles that can be generated by computer from an author's name. Authors of scientific or scholarly publications usually have a view of the bibliographic systems that record their works, but it will come as a revelation to most that these same systems also have a view of them – and not just as "the user" in the abstract, but as specific individuals. In effect, through CAMEOs, databases can deliver them their own personal

thesauri for describing long-term interests. Hardly anyone seems to have exploited this fact. *Grzelak* and *Kowalski* (1983) and *Shepherd* and *Phillips* (1986) automatically profiled individuals on the basis of their publications, but in projects in which that capability was not the main point.

The senses of the word "cameo" in *Webster's New Collegiate Dictionary* include "a small medallion with a profiled head in relief" and "a usually brief literary or filmic piece that brings into delicate or sharp relief the character of a person, place, or event." Thus "CAMEO" seems felicitous in a context where "profiles" and "brief characterizations of persons" take on specialized meanings. At least two other fields, chemistry and education, have also used it as an acronym, but without the wordplay.

Because core-and-scatter distributions of various kinds are commonplace in bibliometrics, it makes sense to distinguish them by the kinds of term used to generate them (input) and the kinds of terms generated (output). The ties of "CAMEO" to concepts of personal identity suggest that this acrononym be limited to distributions generated from a specific author's name. Table 2 shows bibliometric studies that do and do not contain CAMEOs, based on term input. In the language of Dialog, the input would be the argument of a SELECT command, and the output would be the co-occurring values produced by a RANK command. Today, for example, S. C. Bradford might replicate his famous study by entering "SELECT Lubrication/DE" in Ei Compendex to form the set of all articles indexed with the descriptor "Lubrication" and then "RANK JN CONT" to rank journal names in continuous descending order by the number of articles on this topic they contain. That would generate a core-and-scatter distribution – a Bradford distribution no less! – but not a CAMEO.

Table 2
Contrasting bibliometric studies

Studies without CAMEOs	Input	Output	Count used in ranking
Lotka (1926)	Bibliography name	Author names	Works by authors in bibliography
Bradford (1934)	Subject heading	Journal names	Articles with that subject heading in journals
White & McCain (1998)	Journal names	Author names	Articles citing the authors in those journals
Studies with CAMEOs	**Input**	**Output**	**Count used in ranking**
White (2000a)	Author's name	Co-authors	Articles produced with each co-author
White & McCain (2000)	Author's name	Cocitees	Articles cociting input author with others
White (2000b)	Author's name	Citees	Articles by input author citing others

There are various kinds of CAMEOs. Among them are:

(1) Those created from controlled-vocabulary indexing terms applied by others to an author's works, such as descriptors, classification codes, or the LC subject headings of Table 1. (Every author whose works are covered by more than one bibliographic database has multiple CAMEOs in whatever controlled vocabularies they employ. In different databases, exactly the same writings will produce different CAMEOs.)

(2) Those extracted from the natural language of the author's own titles, abstracts, keywords (sometimes called identifiers) or full text. (These natural-language CAMEOs will not differ for the same writings in different databases unless drawn from different fields of the bibliographic record or from nonduplicated sets of identifiers.)

(3) Those ranking the titles or subject categorizations of journals to which an author contributes.

(4) Those created from bylines, ranking an author's co-authors by their frequency of appearance.

(5) Those created from bibliographic references by the focal author or by others. (These will most often be drawn from the citation databases of ISI, the Institute for Scientific Information.)

(6) Those ranking the form classes or genres in which an author publishes.

(7) Those displaying an author's years of publication or ranking them in order of productivity.

The first three kinds of CAMEO prioritize direct indicators of subject matter. The fourth and fifth kinds prioritize the names of authors related to the focal author in various ways. These latter names may be read (by initiates) as indirect indicators of subject matter; they may also designate persons who belong to the author's social network – acquaintances of varying degrees of familiarity. Examples of all these relations, as well as of the last two kinds of CAMEOs, will be given, in somewhat different order, below. Dialog search commands for them will be given as well. The result is a bit of a ramble, but there are many different aspects of CAMEOs to cover.

The "O" in CAMEO can stand for "Offline" as well as "Online." This switch will sometimes be appropriate, since it is possible to download bibliographic records from an online source or a CD-ROM and then produce CAMEOs offline with specialized software such as BibExcel (*Persson*, 2000). The making of CAMEOs is in fact simply one kind of content analysis, a methodology for which offline software is steadily improving.

Two descriptor CAMEOs

While not volatile, CAMEOs are not static either: they can grow and change with the writings from which they are created. An instance would be earlier and later CAMEOs based on someone's growing oeuvre. Another would be the CAMEO created first for one author, and then changed by merging it with a CAMEO created for a second, closely related author. In this sense, CAMEOs are updatable indexing systems of a highly personalized kind. Although no one's long-term interests are fully captured by any set of indexing terms, CAMEOs have more terms than either intermediaries or end-users are likely to produce through introspection or by consulting thesauri. Moreover, as Table 1 revealed, CAMEOs allow useful terms to be *recognized* rather than guessed at or looked up.

The point is well illustrated with CAMEOs made of descriptors. Table 3 shows two that were generated in 1995 from the name of a Drexel professor of electrical engineering, Allen Rothwarf, now deceased. (They are displayed here because, after deletion of singleton descriptors, they can be compared on a single page; their current versions are longer.) The pair comprise descriptors applied at least twice to Rothwarf's journal articles and conference papers. The one from INSPEC is based on 44 documents; the one from Ei Compendex, on 53. The two sets of documents have 37 items in common, and the CAMEOs based on them exhibit obvious commonalities in subject matter – in some cases identical terms (e.g., SILICON, THYRISTORS). An expert in Rothwarf's specialties could doubtless connect other terms in the two lists on grounds of synonymy or other substantive associations. Nevertheless, the many nonduplicate terms illustrate the claim above that different databases produce different controlled-vocabulary profiles. Neither CAMEO can be considered redundant; each by itself offers an abundant conspectus of Rothwarf's interests in the controlled vocabulary of a particular database, and the two jointly characterize his research in even more detail. The characterizations are furthermore adapted to document retrieval, in that descriptors are the guaranteed means of representing concepts to a database.

The "E" in CAMEO is a reminder that characterizations of authors can be *edited* online. In the editorial process, Rothwarf or an intermediary could choose terms from CAMEOs to capture the interests of a given time. A subroutine within RANK allows terms in any combination to be saved as sets for immediate use in searching – for example, one could save the top six terms and then skip down to save the twelfth and the fifteenth terms if such editing produced the profile one wanted. Another subroutine allows terms similar in meaning to be combined and re-ranked.

Table 3
Descriptor CAMEOs from Allen Rothwarf's publications in two electrical engineering databases as of 1995

Rank		INSPEC		Ei COMPENDEX
1	24	SOLAR CELLS	36	SOLAR CELLS
2	16	SEMICONDUCTOR DEVICE MODELS	10	SEMICONDUCTOR DEVICES
3	12	CADMIUM COMPOUNDS	8	SEMICONDUCTING SILICON
4	12	COPPER COMPOUNDS	7	SEMICONDUCTOR MATERIALS
5	12	II-VI SEMICONDUCTORS	6	SILICON
6	10	ELEMENTAL SEMICONDUCTORS	6	THYRISTORS
7	10	SILICON	5	MATHEMATICAL MODELS
8	10	TERNARY SEMICONDUCTORS	5	THIN FILMS
9	9	HIGH-TEMPERATURE SUPERCONDUCTORS	4	COMPUTER SIMULATION
10	9	INDIUM COMPOUNDS	4	HETEROJUNCTIONS
11	8	AMORPHOUS SEMICONDUCTORS	4	PHOTOVOLTAIC CELLS
12	7	BARIUM COMPOUNDS	4	SEMICONDUCTOR DEVICE MANUFACTURE
13	7	YTTRIUM COMPOUNDS	3	AMORPHOUS
14	6	HYDROGEN	3	CADMIUM SULFIDE
15	6	P-N HETEROJUNCTIONS	3	DESIGN
16	6	THYRISTORS	3	DIFFUSION
17	4	SUPERCONDUCTING THIN FILMS	3	ELECTRONIC PROPERTIES
18	3	INFRARED DETECTORS	3	ELECTRONS
19	3	INSULATED GATE FIELD EFFECT TRANSISTORS	3	MEASUREMENTS
20	3	INTERFACE ELECTRON STATES	3	PHOTOCONDUCTIVITY
21	3	SUPERCONDUCTING TRANSITION TEMPERATURE	3	PHOTOVOLTAIC EFFECTS
22	2	ANNEALING	3	SEMICONDUCTING FILMS
23	2	BISMUTH COMPOUNDS	3	SEMICONDUCTOR DEVICES, MOSFET
24	2	CALCIUM COMPOUNDS	3	SILICON ON SAPPHIRE TECHNOLOGY
25	2	CARRIER DENSITY	2	ABSORPTION
26	2	CARRIER LIFETIME	2	ANALYSIS
27	2	CARRIER MOBILITY	2	APPLICATIONS
28	2	DEEP LEVELS	2	BAND STRUCTURE
29	2	DIGITAL SIMULATION	2	CERAMIC MATERIALS
30	2	DOPING PROFILES	2	CHARGE CARRIERS
31	2	EBIC	2	COPPER COMPOUNDS
32	2	ELECTRONIC ENGINEERING COMPUTING	2	DOPING
33	2	ENERGY GAP	2	EFFICIENCY
34	2	INVERSION LAYERS	2	ELECTRIC CURRENTS
35	2	PHOTOCONDUCTIVITY	2	FABRICATION
36	2	SPUTTERED COATINGS	2	HIGH TEMPERATURE SUPERCONDUCTORS
37	2	STOICHIOMETRY	2	JUNCTIONS
38	2	STRONTIUM COMPOUNDS	2	MODELING
39	2	SUPERCONDUCTING ENERGY GAP	2	SEMICONDUCTING CADMIUM COMPOUNDS
40	2	SUPERCONDUCTING JUNCTION DEVICES	2	SEMICONDUCTOR DEVICE MODELS
41	2	SUPERCONDUCTIVE TUNNELLING	2	THEORY
42	2	THICK FILMS	2	TRANSISTORS
43	2	THIN FILM DEVICES	2	TRANSISTORS, FIELD EFFECT
44	2	ZINC COMPOUNDS	2	TUNNELING

These terms are then marshaled in search strategies to find documents. The familiar Boolean operators, AND, OR, and NOT, impose meanings on terms in combination. In 1993, moreover, DIALOG introduced the TARGET command, which dovetails with RANK by permitting retrieval based on lists of terms, such as RANK generates, without the need for relating the terms with Boolean operators. TARGET produces as output a list of up to 50 bibliographic records, ranked as to probability of being relevant on the basis of frequency of occurrence of the input terms within the retrieved set.

A TARGET option complements the editing capabilities within RANK. It allows one to weight terms for retrieval purposes by marking those that must be present if a document is to be retrieved, in contrast to those that *may* or *may not* be present.

Of course, if CAMEOs are being fine-tuned for use as search profiles, numerous factors might affect the editorial process. Rothwarf often had co-authors, and the terms in Table 3 might not all express his own interests as opposed to theirs. Even if every term shown did express them, some terms would doubtless be more important than others. For example, CAMEOs do not discriminate with respect to time (unless specially constructed to do so), and it is likely that Rothwarf, if presented with the CAMEOs, would have chosen terms reflecting newer interests rather than older. Also, terms not on either list might have been added (from controlled vocabulary or natural language) to capture interests wholly new.

This prompts the question as to how to judge CAMEO quality. Which profile in Table 3 indicates Rothwarf"s interests better? In the absence of well-specified purposes, the answer is not determinable. There is no absolutely definitive profile of an author in indexing terms, just as there is no absolutely definitive portrait or snapshot of a person; there are only multiple views. Assuming no mixups of homonymic author-names and no gross mistakes in indexing, either CAMEO seems reasonably good. Both together – or some eclectic combination of terms from both – may be better still, on grounds of comprehensiveness. Rothwarf himself might have drawn different terms from either or both at different times, depending on his projects of the moment. Someone using Rothwarf's name as a quick way of generating many search terms might choose differently yet again. Moreover, the CAMEO judged superior by, say, the author being profiled might not produce a better retrieval when its terms were used in online searches. All one can know for sure is that each CAMEO depicts an author within a particular database.

Intelligent agents

Here we might note the potential linkage of CAMEOs with "intelligent agent" search technologies. Agents are content-bearing, and CAMEOs are one type of content that they might bear. CAMEO-agents might seek not only conventional writings, but the descriptions and reviews of nondiscursive objects in various media (e.g., pictures, sound recordings, software) that are consulted before retrieving the objects themselves.

Strictly speaking, CAMEO-agents would be browsers rather than searchers, in that their use does not presume a known object. Their purpose is simply to rove the electronic wilderness and find something interesting. In this interpretation, *searching* involves looking for something whose identity is known but whose location is not, whereas *browsing* involves looking for something whose general location is known but whose identity is not (cf. *Zoelick*, 1987). Much online searching, so called, actually involves finding a promising location in subject space and then browsing the set of documents there as a final step. In searches that require browsing, an object can be found only because persons can recognize it on the basis of their interests rather than on fully stated objective criteria. As I wrote some years ago (*White*, 1992, p. 75), "Because a target publication can be more or less well described – for example, 'I want a history of minesweeping' – it is possible to delegate searching to other parties and to expect fair agreement on what would be a successful outcome. In contrast, browsing, like bathing, cannot be delegated; it must be done in person, because no one else has your particular bundle of interests – ultimately, your embodied self – for objects to trigger. With browsing, you must let publications 'search you.'"

If, nevertheless, we want to delegate at least the first pass in browsing to CAMEO-agents, we must try to simulate a person's interests as well as possible. In basing the CAMEO-agent on one or more writings by someone, we attempt to simulate the part of a person that browses – the self as a bundle of interests. Presumably, published writings are a tractable representation of an author, and CAMEOs can condense that representation to key features with reasonable fidelity.

CAMEO technology in fact allows a complete outsider – someone who knows only a published author's name – to model that author's interests quite fully. Any librarian or information specialist doing profiles for a selective dissemination of information service could create CAMEOs even before interviewing the profilee. (The profiler would have a strong description of the profilee to use as an opener and might be able to move directly into the editing phase.) Nor is there any technical reason why authors could not create CAMEOs of their own. The chief barrier at present is ignorance. Almost no academic writers know that CAMEOs can be generated for them; online bibliometric capabilities

have outstripped knowledge of them. It does not seem farfetched, however, to imagine the technology spreading.

A CAMEO for browsing

CAMEOs as browsing agents can be demonstrated with one produced in 1993 for Gary W. Strong, then a Drexel faculty colleague of mine whose research interests included computer simulation of visual processes. INSPEC had assigned 20 descriptors to 10 of his publications. The first eight, ordered by Dialog's RANK command, appeared as in Table 4.

Table 4
Partial descriptor CAMEO from INSPEC for Gary W. Strong

Rank No.	Items	Term
1	3	NEURAL NETS
2	2	HUMAN FACTORS
3	2	VISUAL PERCEPTION
4	1	ADAPTIVE SYSTEMS
5	1	ARTIFICIAL INTELLIGENCE
6	1	BRAIN MODELS
7	1	COGNITIVE SYSTEMS
8	1	COMPUTER SCIENCE EDUCATION

Table 5
50 titles in TARGET retrieval based on a CAMEO for Gary W. Strong

1. Looks recognition by adaptive junction
2. Neuro-morphology of biological vision: emulation and generalization of visual receptive fields by fractional discriminant functions
3. Irreversibility and creativity in neurodynamics
4. An algorithm of selective perception and analysis of environmental information by a neural network under the influence of inner afferentation
5. Visual modelling
6. A solution of the figure-ground problem for biological vision
7. A neural network for the processing of optic flow from ego-motion in man and higher mammals
8. A mathematical model for neuronal response properties and modular organization in the motion processing area of the primate cerebral cortex
9. Learning visual coordinate transformations with competition
10. A neural network model of object segmentation and feature binding in visual cortex
11. Modeling human visual object recognition

Table 5 (continued)

12.	Texture discrimination and binding by a modified energy model
13.	Integrating a cognitive model of information processing with a functional neuroanatomical model of prefrontal cortex processing: implications for understanding prefrontal cortex processing
14.	Neural nets for complex scenes understanding: simulation of a visual system with several cortical areas
15.	Neural network modeling of visual recognition
16.	Neural network model for human visual perception of 3D curvilinear motion
17.	Theoretical framework for analysing the behaviour of real and simulated neural networks
18.	Correlated neuronal firing: a clue to the integrative functions of cortex?
19.	Neural Network Dynamics. Proceedings of the Workshop on Complex Dynamics in Neural Networks
20.	Morphological neural networks and image algebra in artificial perception systems
21.	Comparison of the symbolic and connectionist approaches to modelling intelligence
22.	NMDA-based pattern discrimination in a modeled cortical neuron
23.	An adaptive human visual weighted image coding
24.	Empirically derived model of the role of sleep in associative learning and recuperative processes
25.	Human visual perception and cognitive processes, modeling and its application
26.	Exploration of the attractor space of small networks of reciprocally connected processing elements
27.	A retina-like image acquisition system with wide range light adaptation
28.	A neural network model of dynamic form perception: implications of retinal persistence and extra retinal sharpening for the perception of moving boundaries
29.	Neuronal-morphology of biological vision: a basis for machine vision
30.	Common features of neural-network models of high and low level human information processing
31.	Fodor and Pylyshyn on connectionism
32.	Notationality and the information processing mind
33.	Geometrical representation of the combination of visual area functions
34.	Connectionist models of orientation identification
35.	Alignment, scaling, and size effects in discrimination of graphical elements
36.	Perception of oncoming vehicle time-to-arrival
37.	Perceptual skill and the cerebral hemispheres
38.	An empirical evaluation of tools to aid in enroute flight planning
39.	Electroencephalographic correlates of psychological defense
40.	Multi-modal cockpit warnings: pictures, words or both?
41.	Proceedings of the Fourth Australian International Conference on Speech Science and Technology
42.	The Purkinje unit of the cerebellum as a model of a stable neural network
43.	Temporal order, timing, and probability context effects on pattern recognition and categorization in neural networks
44.	A connectionist approach to effects of anxiety and task difficulty on learning
45.	Basins of attraction in disordered networks
46.	Cognition and neural network modelling
47.	Limitations of logical reasoning in neural networks and reasoning by analogy
48.	A distributed model of the representational states in classical conditioning
49.	Computational analysis of the operation of a real neuronal network in the brain: the role of the hippocampus in memory
50.	Semantic transparency, brain monitoring and the definition of hybrid systems

The agent was made by arbitrarily editing the descriptors in Table 4 to include only Neural Nets, Human Factors, Visual Perception, Adaptive Systems, Brain Models, and Cognitive Systems. In the RANK module these six terms were retained as a temporary SearchSave, which was then executed to form six new sets of documents. The new sets were then passed to TARGET, which looked for up to 50 documents with as many co-

occurring terms from among the six as possible. The TARGET command confines its search to the most recent year of additions to the database (in this case 1992–93). The titles of the resulting 50, relevance-ranked by TARGET, are presented in Table 5.

Granted it is hard to judge such output without full bibliographic citations and abstracts, and one would want Strong himself to have the last word on the relevance and novelty of the items retrieved. The point here is simply to convey the quality of the retrieval in brief. Strong's INSPEC descriptors are unchanged as of 2000, and anyone can replicate the Strong retrieval for greater currency and greater bibliographic information if desired. Also, since many *Scientometrics* readers either know authors or are authors themselves, it should be easy enough for them to test the RANK-TARGET process in a favorite database of their own, with CAMEOs of their own devising.

A different CAMEO-agent can be produced for Strong from the identifiers attached to his publications. Unlike descriptors, which are controlled vocabulary assigned by professional indexers, identifiers are noun phrases extracted from the author's own language (authors sometimes supply them as "keywords," especially in conference proceedings). Table 6 displays Strong's list of identifiers in INSPEC. The set of his publications was formed, and then the Dialog command RANK ID CONT DETAIL was given, meaning, in this case, "RANK Strong's IDentifiers in CONTinuous descending order with DETAILed statistics." The latter include the frequency counts on view in Table 6, which are available for standard types of phrase-indexed fields in many Dialog databases. (Unfortunately, RANK cannot be used with word-indexed fields, such as titles and abstracts, in Dialog. This limitation may not hold for other database vendors.)

Strong's union of computer simulation, artificial intelligence, neural networks, and vision is evident in the listing, but one can also see his work in other areas, e.g., human-computer interface design, undergraduate education in information systems, and information retrieval. To break out the different interests commingled in Table 6, someone would of course have to edit the CAMEO, and this would require not only domain knowledge but also awareness of Strong's preferences of the moment.

The two-column display in Table 6 highlights the transition from the extremely common identifiers at top left to the wholly idiosyncratic ones, beginning with Microcomputer Selection Committee, at lower right. Most of the identifiers in the column at right would retrieve *only* Strong's publications if they were sent out to browse as CAMEO-agents in INSPEC. On the other hand, an agent consisting of top-ranked terms in the column at left – for example, Information, Behavior, Pattern, and Microcomputer – would very likely be too general to retrieve anything but a mishmash. In a CAMEO-agent one would probably depend on combinations of terms from the left column to provide reasonably precise retrieval.

Table 6
Gary W. Strong's identifiers ranked by frequency of occurrence in INSPEC

INFORMATION	445676	ILLUSORY CONJUNCTIONS	8
BEHAVIOR	214114	SERIAL PROCESSES	7
PATTERN	150525	VISUAL MOMENTUM	6
MICROCOMPUTER	60639	OBJECT FEATURE EXTRACTION	5
PERCEPTION	20048	HUMAN INFORMATION PROCESSING CAPABILITIES	5
FEATURE EXTRACTION	14682	CAPABILITIES	5
NEURAL NETWORKS	11313	SPATIAL CODE	5
VISUAL PERCEPTION	9897	TIMULUS ARRAY	4
COMPUTER APPLICATIONS	8457	ORGANIZATIONAL TOOL	4
ADAPTIVE SYSTEMS	7405	ATTENTIONAL PROCESS	3
USER INTERFACE	6040	SPATIAL INFERENCING	3
RECALL	4043	COMPUTER TASK	3
CITIES	3404	INTERCOLUMNAR CONNECTIONS	2
INFORMATION PROCESSING	2912	TAG-ASSIGNMENT	2
MENUS	2336	SPREADING-ACTIVATION MECHANISM	2
THESAURUS	1708	INFORMATION SYSTEMS PROGRAM	2
ICONS	1205	DISPLAY NETWORK	2
HUMAN PERCEPTION	405	MICROCOMPUTER SELECTION COMMITTEE	1
PARALLEL PROCESSES	381	END USER RETRIEVAL	1
HUMAN COMPUTER INTERACTION	357	END-USER DATA ENTRY	1
SEMANTIC NETWORK	350	METHOD-BASED TASKS	1
HARDWARE REQUIREMENTS	316	ASSOCIATIVE INTERFERENCE	1
UNDERGRADUATE COURSE	268	CITY LOCATION	1
HUMAN ACTIVITY	159	COMPUTER-HUMAN INTERFACE DESIGN PRINCIPLE	1
HUMAN INFORMATION	148	COMMAND CONTROL SCHEMES	1
SOFTWARE INTERFACES	142	END-USER INDEXING	1
SELECTIVE ATTENTION	125	FACETED HIERARCHICAL THESAURUS ORGANIZATION	1
SPATIAL PROCESSING	125	NAVIGATIONAL AID COMBINATIONS	1
INFORMATION SPACE	92	POOLED FEATURES	1
INFORMATION NAVIGATION	80	NONSPATIAL PROCESSING	1
HEBBIAN LEARNING RULE	73	CORTICAL PROCESSING SYSTEM	1
TASK STRUCTURE	67	ECONOMICAL INFORMATION RETRIEVAL SYSTEM	1
DISTRIBUTED REPRESENTATIONS	67	OUTCOME-BASED TASKS	1
HODGKIN-HUXLEY MODEL	54	PARALLEL CONNECTIONIST MODEL	1
DREXEL UNIVERSITY	54	SPATIAL SELECTIONAL REQUIREMENTS	1
SPATIAL INDEXING	47	SPATIAL STRUCTURE MAPPING	1
COGNITIVE MAP	41	FACTORED ARCHITECTURE	1
NAVIGATIONAL AID	40	INFORMATION ASSEMBLY LINE	1
MICROCOMPUTER SELECTION	35	TACHISTOSCOPIC RECOGNITION TASKS	1
INTERFACE TECHNIQUES	25	AFFORDANCE-BASED INFERENCE	1
CONTENT INFORMATION	20	TASK-DEPENDENT INTERFERENCE	1
CELL ASSEMBLIES	18	NEURALLY PLAUSIBLE SERIAL MECHANISM	1
COGNITIVE NEUROSCIENCE	17	MECHANISM	1
HUMAN COMPUTER INTERFACE DESIGN	15	TWO-DIMENSIONAL STIMULUS SPACE	1
HUMAN OBJECT RECOGNITION	14	SPATIAL IMAGE COMPUTATIONS	1
DATABASE TOOL	9	INTRACOLUMNAR ACTIVITY	1
MINICOLUMNS	8	LOCATION-RELATIVE POINT-OF-VIEW	1

A sophisticated retrievalist might reweight terms algorithmically, based on their frequencies in Strong's works and in the database (*Salton*, 1975). These frequencies are given in the DETAILed statistics of the RANK module.

CAMEOs and personal networks

In *White* (2000, 2001) and *White* and *McCain* (2000) I have named and analyzed three important kinds of CAMEOs that are uniquely available from Scisearch, Social Scisearch, and Arts & Humanities Search, the online citation databases of the Institute for Scientific Information (ISI). These are:

(1) the *citation identity,* which consists of all authors cited by a focal author;

(2) the *citation image-makers,* consisting of all authors who cite a focal author and thereby create his or her citation image;

(3) the *citation image,* which consists of all authors cocited with a focal author.

Another kind also consists of personal names, but is not unique to ISI:

(4) the focal author's *co-authors.*

Table 7 shows where these four profiles fit in a breakdown of CAMEO types. The column divider is whether the CAMEO has been generated from text that is the focal author's own or from related texts produced by others. The row divider is whether a given CAMEO can be called up from bibliographic databases in general (including those of ISI) or only from ISI databases.

The content of items in the left column of Table 7 is governed mainly, if not entirely, by authorial choices (e.g., of co-authors, genres, journals submitted to, vocabulary used). The citation identity is then simply one more feature emerging from a focal author's own texts as cumulated in the oeuvre. In contrast, the citation image and the list of image-makers can be taken as indexing of the focal author's work by other publishing authors. The latter thus join forces with professional indexers, catalogers, and classifiers, but bring their own subject expertise to bear on what they cite.

The co-author, identity, image, and image-maker profiles can be interpreted as simultaneously representing both subject matter and ties to persons. Since the latter can be rendered as personal networks, CAMEOs can provide data for ego-centered network analysis (*Wasserman* and *Faust*, 1994). Authors, that is, may be graphed as a network of linked nodes, with the links standing for explicit relationships centered on the focal author ("ego"). In social-network terminology, *undirected* links imply joint memberships, such as occur in the focal author's cocitation image or set of co-authors.

Table 7
Some types of CAMEOs

	Data from Author	Data from Others
All Databases	Co-author Journal Title Abstract Full-text Keyword Identifier Genre Publication year	Descriptor Subject heading Classification code
ISI Databases	Citation identity	Citation image-makers Citation image

They are counted as *degrees* of his or her node and graphed as lines. In contrast, *directed* links imply one-way choices, graphed as one-way arrows. They are exemplified by the focal author's image-makers or identity. The image-makers who cite a focal author are counted as *indegrees* (arrows in) to his or her node. The focal author's own citees in the identity are counted as *outdegrees* (arrows out) from it. We could analogously speak of *incitations* and *outcitations;* some make the same distinction by contrasting *citations* with *references.*

The matter is complicated in that these relationships are not simply present or absent and hence codable as ones or zeroes. Like subject terms, they can acquire weights, based on the number of times the relation occurs. Over time, for example, a focal author might have co-authored papers with the same colleague six times, cited a second person twice, been cited by a third 10 times, or cocited with a fourth 100 times. The links in any resulting graph are thus termed *valued* as opposed to one-and-zero binary. In common with other network analysts, several bibliometricians have rendered valued graphs by varying the thickness of the links to suggest the weights and pruning out links whose weights fail to meet a certain threshold. The graphs in *White* and *McCain* (2000) and *White* (2000) are both pruned to include only the top 50 cocitees from image CAMEOs for Belver Griffith and Eugene Garfield respectively. For many thousands of authors, weighted links with other authors are so numerous that they cannot be shown in standard graphical displays, and so a pruning rule of some sort is necessary.

The links in author-centered citation networks all reflect intellectual ties; links that coincide with personal acquaintanceships reflect social ties as well. Possible interactions between social ties and citation choices have been studied since the late 1970s (*Mullins*

et al., 1977; *Lenoir*, 1979). While nothing is yet settled, the evidence suggests that social ties have some influence, but ties of perceived relevance have more (*Stewart*, 1990; *Peters* et al., 1995; *Vinkler*, 1998; *Baldi*, 1998; *White*, 2001). The mild claim here is that an author's identity, image-makers, or image – all easily obtained – can assist in clarifying the matter further.

While co-authors are almost always personal acquaintances, persons both known and not known to the focal author may appear in citation-based CAMEOs. In the identity, for example, a citer's complete network of citees can range from someone as close as a spouse to obscure historical figures no one living could have known. The identity thus seems worth trying as a data-gathering probe, one that can be used to question a citer on whether social ties with citees exist and, if so, on their nature. The image or image-maker names could similarly be used as probes.

Both *Freeman* (1984) and *White* (2001) list categories of personal acquaintanceship that maight be used with CAMEOs in science studies. Freeman's discussion includes, for example, a scale originally used by Barry Wellman in a 1977 survey of their fellow social-networks researchers: (1) familiar with, (2) follow the work of, (3) know, (4) teacher of, (5) student of, (6) colleague of, (7) attended conference with, (8) (ever) co-worker with, (9) co-author with. The first two relations imply mere intellectual ties; the other seven are also social. A pioneering study by *Murray* and *Poolman* (1982) relates ties of various strengths to the ways in which scientists find and cite documents.

Griffith as example

Table 8 illustrates the CAMEO types of the preceding section. They comprise author-names variously connected to Belver C. Griffith's in the combined ISI databases in August 2000. Within ISI's limits of coverage, Griffith's co-author CAMEO is complete across the 36 publications of which he was sole or first author; the others have been cut at arbitrary points to create a manageable table. (Griffith's co-author CAMEO has been emended to correct an entry in the ISI database that fails to list his co-authors in a 1974 article.) The ISI files do not go back far enough to cover his very productive years before 1974, nor do they cover his publications in media other than scholarly journals (e.g., his technical reports and book chapters). The Table 8 CAMEOs are therefore offered here for purposes of demonstration rather than as a complete bibliographic history.

Table 8
CAMEOs for Belver C. Griffith from ISI data

	Co-Authors		Citation Identity		Image-Makers		Citation Image
36	GRIFFITH BC	11	GRIFFITH BC	17	McCAIN KW	368	GRIFFITH BC
7	DROTT MC	9	PRICE DJD	15	GARFIELD E	207	SMALL HG
6	WHITE HD	9	SMALL HG	15	GRIFFITH BC	147	PRICE DJD
2	CHERNYI AI	4	GARVEY WD	13	SMALL HG	120	GARFIELD E
2	GILYAREVSKII RS	3	BROOKES BC	12	WHITE HD	93	CRANE D
2	MARKUSOVA VA	3	MARSHAKOVA IV	10	VLACHY J	80	KUHN TS
2	McCAIN KW	3	MEADOWS AJ	9	HARGENS LL	72	MULLINS NC
2	SAYE JD	3	SARACEVIC T	7	MULLINS NC	69	MERTON RK
2	SERVI PN	3	VIRGO JA	5	CHUBIN DE	62	HAGSTROM WO
1	ALLISON PD	2	COLE JR	5	COLLINS R	57	COLE JR
1	ANKER AL	2	CRANE D	4	BLACKMAN LS	56	COLE S
1	BEARMAN TC	2	FEYERABEND PK	4	BURGER AL	48	LINE MB
1	DEY S	2	GARFIELD E	4	CAWKELL AE	46	GARVEY WD
1	GARVEY WD	2	GREEN PE	4	MORAVCSIK MJ	45	MULKAY MJ
1	MANCALL JC	2	JOHNSON SC	4	SCHUBERT A	45	NARIN F
1	MARKUSOVA V	2	KORTEN F	4	SMITH LC	41	BROOKES BC
1	MORAVCSIK MJ	2	KRUSKAL JB	3	ARABIE P	41	CHUBIN DE
1	PRICE DJD	2	KUHN TS	3	BOORMAN SA	41	MORAVCSIK MJ
1	SMALL HG	2	MITROFF II	3	DROTT MC	41	WHITE HD
1	STEWART JA	2	MULLINS NC	3	EGGHE L	38	GILBERT GN
1	STONEHILL JA	2	ORR R	3	LANCASTER FW	37	ZUCKERMAN H
		2	SANDISON A	3	LEYDESDORFF L	36	McCAIN KW
		2	SHEPARD RN	3	MARTON J	31	BENDAVID J
		2	URQUHART JA	3	MOED HF	31	SULLIVAN D
		2	ZUCKERMAN H	3	RAO IKR	29	EDGE D
				3	SPITZ HH	28	GASTON J
				3	SULLIVAN D	28	KAPLAN N
				3	WHITE DH		

As many have noted, unintentional inconsistencies in citing and data-entry practices fill ISI databases with variant names for the same entity; in *White* (2001) I dubbed these variants *allonyms.* In Table 8, allonyms of author-names have been gathered and their counts combined under a single name-form (e.g., "Price DJD"). This may occasionally produce double-counting when a single author is cited in two different ways in the same article (e.g., as "Small H" and "Small HG"). Moreover, to approximate as closely as possible Griffith's own taste in citees, I based his citation identity only on writings of which he was the sole or first author. This was done because the raw data from ISI on Dialog conflate his personal citees with those of some of his co-authors when the latter led the byline – M. Carl Drott, myself, and so on. A discrepancy results: in Griffith's pared-down identity, he cites himself in 11 items, whereas, among his own image-

makers, he cites himself in 15 items. The latter count is higher because the image-maker CAMEO automatically included four items citing him that were by others as first author and him as a secondary author; these conflations were not edited out.

The coverage and conflation problems mentioned here will affect all CAMEOs like those in Table 8, and conscientious researchers will want to be aware of them as they proceed.

Turning to content, the CAMEOs in Table 8 are not a mere jumble of names; they exhibit various kinds of structure. Although all are core-and-scatter distributions, the one having the weightiest co-occurrence counts is the citation image, making it the most robust depiction of an author's intellectual world (*White*, 2000). Formed by the aggregate perceptions of hundreds of citers, the image is a consensual view of authors strongly related over time. Substantively, it is appropriate that Griffith's three top cocitees are Henry Small, Derek Price, and Eugene Garfield, and that the remainder of the list comprises many bibliometricians, citation analysts, and sociologists of science (cf. the map based on Griffith's cocitation image in *White* and *McCain*, 2000).

On the social front, Griffith was venturesome in meeting fellow researchers throughout his career and knew most of the high-ranking authors in his image on a first-name basis (McCain and I worked in the same building with him; Small and Garfield at ISI and Crane at the University of Pennsylvania were only short walks away). This anecdotally supports use of the image CAMEO as a probe in studies that combine bibliometrics with sociometrics. It may, however, be sociometrically cleaner to use the image-maker and identity CAMEOs as probes, in that they reflect choices attributable to individual citers rather than citers in aggregates.

An obvious structural question arising from Table 8 is whether any automatically produced name appears in all four CAMEOs, which would seem to be an indicator of that person's importance to the profilee. In Griffith's case, Small's name does. The indicator thus works well here; there can be no question that Small and Griffith, who were friends for many years, also loomed large in each other's professional lives, not least because Griffith showed how to extend Small's invention of cocitation analysis to the mapping of science. If Griffith's CAMEOs went back far enough in time, the sociologist Nicholas Mullins would also appear in all four. Mullins, too, was a highly significant figure for Griffith; for example, they published an article in *Science* together, and Griffith provided Mullins with data for an early study of congruence between cocitation and social ties (*Griffith* and *Mullins*, 1972; *Mullins* et al., 1977). Anyone similarly ubiquitous in another author's CAMEOs should be similarly important.

The principle covers multiple appearances in general. For example, Garfield turns up in three of Griffith's four CAMEOs, which will not surprise readers of *Scientometrics.* But beyond a few well-known names, the number of cross-CAMEO appearances falls off quickly, and, to examine social structure further, we must look at *who chooses Griffith* when he is cited by his image-makers and *who is chosen by him* when he cites others in his identity. Raising the matter of direction in citing here may illuminate it for all ISI-based CAMEOs. My guess is that, in the directed relationships, citation is affected by a sense of intellectual seniority. Stated baldly, there is an apparent tendency to cite *up* or *across* the seniority chain much more than *down* (cf. *White*, 2000).

A key example in Table 8 is Derek Price. Griffith in his identity frequently cites Price as an influence. Griffith's image also reveals the bond with Price; 147 cociting articles register a pervasive sense of their intellectual kinship. But Price, signally, is not high among Griffith's image-makers. The reason is not that Price did not respect Griffith's work; much evidence can be produced to the contrary, and the two were also close personal friends. Rather, Price was older and became a prominent author, with an international audience, earlier than Griffith. In that sense Price was intellectually senior to him, as I think Griffith would be the first to admit. By hypothesis, the flow of citation in such cases will be asymmetrical, with more going from junior to senior than the reverse. As one of thousands of possible examples, Griffith cites Price much more than Price cites Griffith.

Call it a matter of intellectual cohorts. One may hesitate to define cohorts precisely in terms of birthdates, but most scholars and scientists have a keen sense of the authors in their fields who arrived, in the reputational sense, some years before they themselves did – the seniors. Scholars and scientists also know persons in their own cohorts whose reputations were made concurrently with their own – their peers. The final grouping – their juniors – consists of persons who were coming up when their own reputations were already made.

Thomas S. Kuhn and Robert K. Merton, both of whom Griffith knew, would for him be seniors like Price. Griffith's peers would include such co-authors as William Garvey, Henry Small, and Nicholas Mullins; also colleagues like F. W. Lancaster and B. C. Brookes who rose with him in the 1960s and 1970s. His juniors would include Carl Drott, Kate McCain, and me, his younger co-authors at Drexel. It will be seen in Griffith's identity that he recurrently cites members of the first two groups but not of the third. Drott, McCain, and I, in other words, turn up among his image-makers in Table 8 (sometimes citing him incidentally as our co-author when we cite ourselves), but we do not appear in his identity there; the same asymmetry noted between Price and him

extends to Griffith's relationship with us. It is tempting to add further personal examples, but perhaps readers should try my claims with CAMEOs of their own choosing, and someone should test the hypothesis statistically. Direction of citation is not ironclad, of course, but if seniors tend to cite juniors less than the other way round, then interest is heightened when a junior ranks high in some senior's identity (*White*, 2000).

There may now be a few additional things to look for when bibliometric data are used to explore social networks. I should add that there is no call for cynicism here, no implication of snobbery or toadying or manipulative practice. In *White* (2001) I argued that a law of least effort governs much citation behavior. Such a law would be consistent with repeated references over time to a relatively small group of seniors and peers who were influential in the earlier part of one's career. The simplest explanation is that, after a certain point, one recycles this select group of orienting figures, many of whom may be personal acquaintances, to conserve reading effort. If that is a vice, it is a nearly universal one. I also argued there that the main reasons for citing are perceived relevance and rhetorical usefulness. In service to both, anyone could reasonably claim that citing established authors over newcomers is a sign of informed respect, not of a vulgar desire to impress. This is not to imply that citer motivations are never questionable, merely that it is wrong to view the entire citation process as amusingly corrupt. The satirical view (*Thorne*, 1977; *Latour*, 1987) relies on credulous cynicism in readers; as theory, it requires very complicated explanations of everything – almost a new anecdote for every citation. That seems reason enough to reject it for something more even-handed. CAMEOs, in any case, can be brought to bear on the question. The reader might look again at Griffith's identity and image-makers to see whether grounds for cynicism appear; I confess I can see none.

The four types of CAMEOs on view in Table 8 are potentially useful as browsing agents, just like the ones based on overtly topical terms that were seen earlier. Various combinations of names can be chosen for ANDing together or to be entered as blocks into TARGET. These forms of retrieval require literacy in a research area and a certain sophistication in online searching; one has to know, for example, whether one wants works *by* the selected authors or *citing* them, and then how to convey that to the system. Possibly the easiest CAMEO to work with when author-names are conjoined to imply subject matter is the citation image (*White*, 1986).

Table 9
Publication CAMEOs for Belver C. Griffith

	Journals
7	Journal of the American Society for Information Science
6	Scientometrics
3	ASLIB Proceedings
3	Journal of Documentation
2	Information Processing & Management
2	Library Quarterly
2	Proceedings of the American Society for Information Science
1	Behavioral and Brain Sciences
1	Bulletin of the Medical Library Association
1	Communication Research
1	Contemporary Psychology
1	IEEE Transactions on Professional Communication
1	Journal of Classification
1	Physics Today
1	Publishing Research Quarterly
1	Science Studies
1	Science, Technology & Human Values
1	Social Studies of Science

	Year
1	1974
2	1975
3	1976
2	1978
4	1979
2	1980
3	1981
3	1982
1	1984
2	1986
3	1987
3	1988
1	1989
2	1991
1	1994
2	1996
1	1997

	Journal Subject Codes
26	Information Science & Library Science
2	Communication
2	Psychology
1	Computer Science, Interdisciplinary Applications
1	History & Philosophy of Science
1	Neurosciences
1	Physics
1	Social Issues
1	Social Sciences, Mathematical Methods

	Document Types
23	Article
6	Book Review
4	Letter
1	Editorial
1	Note
1	Reprint

The four additional CAMEOs for Griffith in Table 9 demonstrate how ISI data can be used to capture other aspects of a publication career. They are based on the same set of publications by Griffith and his co-authors that were used in Table 8. It will be recalled that these 36 items are by no means Griffith's complete output, but they comprise a substantial part of it during 1974–1997. The upper CAMEO lists the frequency with which he published in various journals, *JASIS* and *Scientometrics* being the top two. The CAMEO at bottom left gives ISI's record of his output by year; even this partial list shows his sustained productivity. (To list Publication Years early to late as in Table 9, the Dialog command is RANK PY CONT ALPHA. Without ALPHA the same command orders the years high to low by documents produced; in Griffith's case that puts 1979, with four documents, on top.) The CAMEO headed Journal Subject Codes shows how ISI classifies the journals in which Griffith published (not his articles themselves); he was, of course, strongly but not exclusively identified with information science (26 of 36 items). Among Document Types, which also may be called form classes or genres, he published mostly articles (23 of 36 items); his six book reviews included three of different books by the same author, A. J. Meadows writing on communication in science.

When someone is studying an author's oeuvre for its own sake, CAMEOs like those in Table 9 can be rapidly produced as documentary evidence. They reveal at a glance the journals in which an author has published, interdisciplinary breadth as shown by the Subject Codes, the year-by-year record of output, and favored forms of discourse. They might also help to make a case when an academic author applies for tenure or promotion, since they highlight features of a career that would otherwise be buried in the list of publications attached to a vita. It is easy to produce visualizations of CAMEO counts in, e.g., Excel or DeltaGraph.

Online search commands

CAMEOs can be understood as a form of data mining in bibliographic databases. Table 10 contains the Dialog field tags used to create them; they are all two-letter designators such as AU for the author field and CA for the cited author field in bibliographic records. By varying the tags after SELECT and after RANK as listed in Table 10, the various kinds of CAMEOs found in the present article can be made (other CAMEOs than those seen here are possible).

Any tag under SELECT in Table 10 can be given an argument and used to form a set. The argument must be an author's name, or combined allonyms of the name, in a form appropriate to the database. The appropriate version(s) of an author's name may

have to be looked up in an online index with an EXPAND command. Once entered, SELECT will form the set of writings by that author (AU), or that cite that author (CA), or about that author (DE). RANK can then be entered, followed by any of the tags under it in Table 10. This will call up and rank (high to low) the terms in the set with that tag. If the RANK command is entered right after the set has been formed, it will operate on it by default; if RANK occurs later, it must contain the number of the set on which it is to operate.

Table 10
Field tags for making CAMEOs

ISI or other databases	Select	Rank	Types of CAMEOs
	AU	AU	Co-authors
	AU	PY	Publication Years
	AU	CL	Classifications
	AU	DE	Descriptors
	AU	ID	Identifiers
	AU	JN	Journals
	AU	DT	Document Types
	DE (author's name)	DE	Descriptors
ISI databases	Select	Rank	Types of CAMEOs
	AU	SC	Journal Subject Codes
	AU	CA	Citation Identity
	CA	CA	Citation Image
	CA	AU	Citation Image-makers

In the following example, ? is the Dialog prompt and Griffith's name is used illustratively in the form suitable to ISI databases (no comma after the surname, no spelling out of given names, no periods after the initials):

? SELECT AU=GRIFFITH BC
? RANK CA CONT

Other examples of the same template have appeared earlier; this combination in an ISI citation database would yield Griffith's citation identity. The addition of CONT assures that the searcher will be asked to declare how many author-names are to be ranked; 50 is the default. Only 8 names will be ranked by default if CONT is not used.

Not all databases support all the tags listed in Table 10 (for example, the ID and DE fields have only recently been added to ISI databases, and the CL field does not appear in them at all). Therefore, exploration or expert advice may be needed when databases are unfamiliar. If in doubt as to which database to choose, one can give an author's

name to Dialindex, the online index to Dialog databases, and learn the databases in which non-empty sets are formed. Each of these databases will yield one or more CAMEOs.

Table 11
CAMEO made of subject headings from works about William Morris

Rank	Count		Rank	Count	
1	195	MORRIS, WILLIAM	26	8	RUSKIN, JOHN
2	193	1834-1896	27	8	1819-1900
3	92	19TH CENTURY	28	7	POETIC WORKS
4	84	GREAT BRITAIN	29	5	HOMES AND HAUNTS
5	52	HISTORY	30	5	LITERATURE
6	51	BIOGRAPHY	31	5	MEDIEVALISM
7	46	DESIGNERS	32	5	SIR
8	42	AUTHORS, ENGLISH	33	4	ART
9	41	SOCIALISTS	34	4	ENGLISH POETRY
10	35	CRITICISM AND INTERPRETATION	35	4	FINE EDITIONS
11	30	ENGLAND	36	4	LITERARY LANDMARKS
12	27	KNOWLEDGE	37	4	LITERATURE AND SOCIETY
13	20	PRINTING	38	4	LITERATURE, COMPARATIVE
14	18	CONTRIBUTIONS IN DECORATIVE ARTS	39	4	MANUSCRIPTS
15	18	SOCIALISM	40	4	MORRIS & CO. (LONDON, ENGLAND)
16	16	ARTS AND CRAFTS MOVEMENT	41	4	PSYCHOLOGY
17	16	EXHIBITIONS	42	4	SAGAS
18	16	POLITICAL AND SOCIAL VIEWS	43	4	SHAW, BERNARD
19	15	HISTORY AND CRITICISM	44	4	SOCIALISM IN GREAT BRITAIN
20	14	BOOK DESIGN	45	4	YEATS, W. B
21	13	BIBLIOGRAPHY	46	4	1856-1950
22	13	KELMSCOTT PRESS	47	4	1865-1939
23	11	DECORATIVE ARTS	48	3	APPRECIATION
24	11	INFLUENCE	49	3	BURNE-JONES, EDWARD COLEY
25	8	FRIENDS AND ASSOCIATES	50	3	CATALOGS

There is a CAMEO type in Table 10 that has not yet been illustrated – the one in which an author's name follows SELECT as a descriptor (DE) – that is, as a heading for writings *about* that author. If one forms the set of these writings and then asks for a ranking of *their* descriptors, writings that are numerous and variegated may yield an interesting profile. Table 11 gives the top 50 subject headings for the Victorian genius William Morris. They were taken from the five REMARC databases and LC MARC–Books, which run in their coverage from the nineteenth century to the present. Morris's many-splendored life is refracted through noun phrases, of which the CAMEO in full has150 more.

Mapping an image CAMEO

As noted above, *White* and *McCain* (2000) maps authors heavily cocited with Belver Griffith; *White* (2000) does the same for Eugene Garfield. Like the maps just mentioned, Figure 1 converts a citation image CAMEO into a PFNET or Pathfinder network (*Schvaneveldt*, 1990). Rather than re-mapping Griffith, I introduced his major influence Derek John de Solla Price as the focal author in Figure 1. Price's image here was drawn from a special database, a 10-year run of the *Arts & Humanities Citation Index* (AHCI 1988–97), which the Institute for Scientific Information gave my college at Drexel as a research grant in 1998.

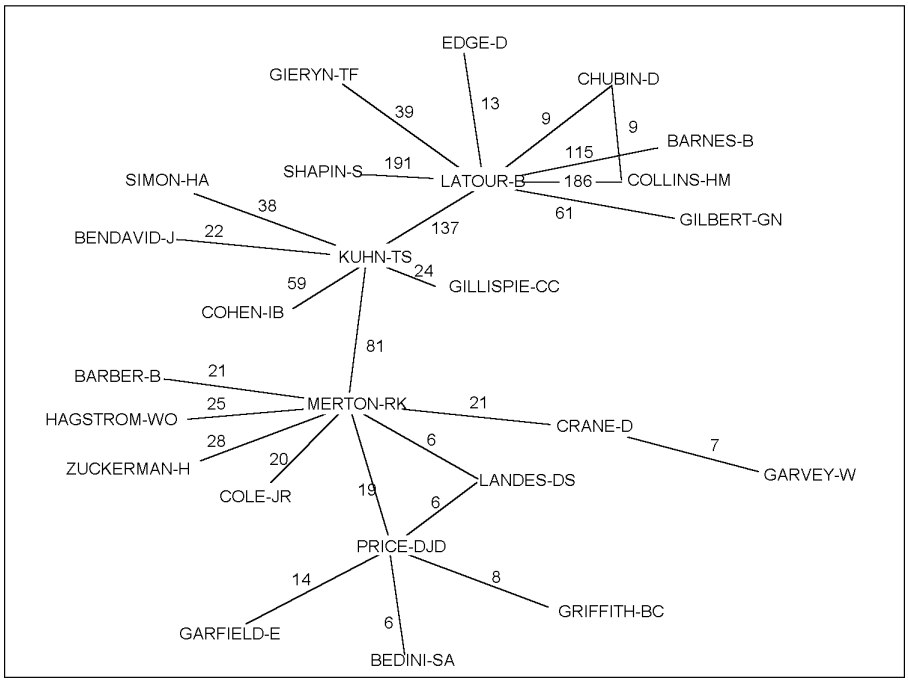

Figure 1. PFNET of 24 authors in Derek J. de Solla Price's cocitation image

The mapping software is part of a new system called AuthorLink that Xia Lin, Jan Buzydlowski, and I developed at Drexel during 1998–2000. It runs on top of BRS/Search software, a commerical package (*White* et al., 2000). To minimize the user's cognitive load, it is designed to require only a single name as input (e.g., PRICE-DJD, with surname and initials hyphenated). When one is entered, AuthorLink matches it against an index file drawn from the 10-year AHCI. The BRS command TALLY (like Dialog's RANK) produces a rank-ordered list of the 24 authors most cocited with the focal author. AuthorLink then systematically pairs all authors in the list, finds the cocitation counts for all pairs, and places them in a 25-by-25 upperhalf matrix. Theoretically, the higher the counts, the more similar the cocitees. From the matrix AuthorLink generates Kohonen feature maps and PFNETs at user's request. All of this is done in real time; the wait from entering the focal author's name to viewing of the maps is generally no more than 10 seconds.

PFNETs usefully simplify the information in the 25-by-25 matrix. The Pathfinder algorithm regards the cocitation counts shared by any two authors as weights. It requires that two parameters, r and q, be set; as *Chen* (1998) explains, r defines the Minkowski metric used to compute the distance of a path, and q constrains the scope of minimum-weight paths that are considered. When, as in AuthorLink, r = infinity and q = number of authors minus one, the algorithm produces a graph in which all author-nodes are connected, but only with paths that do not violate the triangle inequality – that is, with paths whose weights are less than the summed weights of any other paths in combination. The effect is to prune all paths except those with the single highest (or tied highest) cocitation counts between authors.

These remaining "minimum cost" paths are usually quite meaningful substantively. For example, Price is directly linked to his fellow bibliometricians Belver Griffith and Eugene Garfield because they are more highly cocited with him than any other author. This should make substantive sense to readers of this journal, and indeed the direct Price-Griffith and Price-Garfield paths also appeared in the maps in *White* and *McCain* (2000) and *White* (2000), which were created outside AuthorLink. *Jarneving* (1999), a study of the authors most highly cocited in *Scientometrics*, draws its data from a different database, Social Scisearch on Dialog, but independently confirms the same strong linkages. In contrast, Herbert A. Simon in this dataset is more highly cocited with Thomas S. Kuhn than with Price or anyone else, and so only the Kuhn-Simon path is drawn. (Although Price has at least six cocitations with every other author in Figure 1, a few authors cocited six times with him were cut from the map because of AuthorLink's arbitrary limit of 25 names, imposed for the sake of speedy mapping.)

Price's direct links to the nonbibliometricians S. A. Bedini, David S. Landes, and Robert K. Merton reflect his work as a historian of science. He also published accounts of early scientific instruments, which leads to a distinction no other bibliometrician can claim: under the allonym PRICE DJ, he is recurrently cocited in AHCI with Chaucer, who wrote a treatise on the astrolabe in addition to his poetry.

Figure 1 is in fact based on only one of the image CAMEOs that could be generated for Price, whether in Drexel's 10-year AHCI or in any of the full-scale ISI citation databases. His name has so many allonyms in ISI data that he himself joked about it; his ISI-style surname appears not only as PRICE but as DESOLLAPRICE and SOLLAPRICE, and these variants can appear with different forms of his initials. In Drexel's AHCI, the allonym PRICE DJD may be the most common, occurring 86 times; in any case it is obviously picking up valid cocitees.

Figure 1 shows a central Merton-Kuhn axis with offshoots in history of science (Simon, Landes, Cohen, and Gillespie) and American sociology of science (Ben-David, Zuckerman, Barber, Hagstrom, Crane, Cole, and Griffith's old partner Garvey). Price is here seen as an adjunct to Merton; that is, standard-model bibliometrics, as represented by Price, Garfield, and Griffith, appears as an adjunct to traditional American sociology of science. Counter-Mertonian science studies are anchored in Latour, who is seen as radiating from Kuhn. Readers may recall that Latour's constructivist account of citation has been proposed as an alternative to Merton's (*Luukkonen*, 1997), and that Edge, Collins, Gilbert, and Chubin, who are linked to Latour, have offered various critiques of bibliometrics and citation studies.

The numerals on the paths are the cocitation counts for author-pairs, which AuthorLink can optionally display. This being a picture of Price and American sociology of science from the perspective of citers in the humanities, the counts for Latour and Kuhn, and for Latour and other constructivists (Barnes, Shapin, Collins), are an order of magnitude higher than any counts in the lower part of the map. A map based on Price's image CAMEO from Scisearch and Social Scisearch would have some of the same names, but much higher cocitation counts for the "Merton-Price-Garfield" group.

Even so, this PFNET of Price's image CAMEO, like the earlier ones for Griffith and Garfield, is remarkable in what it conjures up from a single name. It highlights major alliances and divisions of an intellectual world in a way that captures insiders' knowledge, yet it can be produced for outsiders through algorithmic means.

The personalization issue

CAMEOs grow out of that part of bibliometrics that, starting with *Lotka* (1926), has been concerned with the papers and oeuvres of particular authors. In this tradition, the citation analyses of Derek Price, Henry Small, and Belver Griffith stand as milestones, and the growing body of work on author cocitation analysis carries forward their concern with papers and oeuvres by specific individuals as units of analysis (*White* and *McCain*, 1989; 1997; 1998). Long before Orbit's GET or Dialog's RANK actually appeared, I called for a command that would "Bradfordize" data online (*White*, 1981), and even then I looked forward to Bradfordizing data from authors' oeuvres, such as their topical or citation histories. That is exactly what CAMEOs do, and I think Griffith, my partner in introducing author cocitation analysis, would share my fascination with them.

The question may arise as to whether they are scientometric in nature. CAMEOs personalize, and some might argue that scientometrics, like science, must be impersonal. Scientometricians, in any case, tend to focus less on authors than on journals, organizations, specialties, disciplines, or nations. Yet all of these macro-units can be translated downward, either in principle or in fact, into the authors that constitute them. The same holds true for micro-units such as books, articles, patents, or even individual knowledge-claims: they can be translated upward into authors. The author-level is where science instersects with softer fields such as biography and scientific communication, or history and sociology. One might go so far as to say that the author-level is where, in the journalistic sense, science gains human interest. Thus, a scientometrics that incorporates CAMEOs is likely to be more interesting than one that does not, even if depersonalized writings are judged more important for the field.

One should also bear in mind that CAMEOs are not merely qualitative. They incorporate measures – term weights – that can be used for algorithmic discriminations in cocited author mapping, social networks research, document retrieval, and the study of professional careers. What is being measured across all these activities is authors' interests, a phrase that connotes not only intellectual attractions, seen in rankings of topics, but also interpersonal ties, seen in rankings of authors. CAMEOs are thus sociocognitive measures, with potential for making scientometrics both more sociological (through studies of networks of authors, which may include "interest groups" of sorts) and more psychological (through studies of authors as they prioritize topics and genres over time). My hope is that some readers will find CAMEOs a seductive new form of data for scientometric analyses.

Probably the most immediate testing ground for CAMEOs, however, is document retrieval. Here, the CAMEO initiative is distinguished from traditional research by its strong personal grounding. Much experimental work in retrieval has involved hypothetical users who present the system with hypothetical queries. CAMEO retrieval, by contrast, derives from the writings of real people with definite interests and particular work to do. The means are now available to profile them both more quickly and in greater depth than ever before. These profilees not only have needs, they have bylines, and bylines are the beginning of CAMEOs. It remains to be seen how good, in their eyes, are the results.

References

BALDI, S. (1998), Normative versus social constructivist processes in the allocation of citations: A network-analytic model, *American Sociological Review*, 63: 829–846.

BJORNER, S. N. (1990), Overcoming the fear of GETting: ORBIT's GET command for first-time users, *Online*, 14: 90–93.

BRADFORD, S. C. (1934), Sources of information on specific subjects, *Engineering*, 137: 85–86.

CHEN, CHAOMEI (1998), Generalised similarity analysis and pathfinder network scaling, *Interacting with Computers*, 10: 107–128.

CHRISTENSEN, F. H., P. INGWERSEN (1996), Online citation analysis: A methodological approach, *Scientometrics*, 37: 39–62.

FREEMAN, L. C. (1984), The impact of computer based communication on the social structure of an emerging scientific specialty, *Social Networks*, 6: 201–221.

GRZELAK, H., K. KOWALSKI (1983), Automatic construction of information queries, *Information Processing and Management*, 19: 381–389.

GRIFFITH, B. C., N. C. MULLINS (1972), Highly coherent groups in scientific change, *Science*, 177: 959–964.

HUDNUT, S. K. (1993), Finding answers by the numbers: statistical analysis of online search results, *Proceedings, 14th National Online Meeting*, New York, May 4–6, 1993. Medford, NJ: Learned Information. pp. 209–219.

INGWERSEN, P., F. H. CHRISTENSEN (1997), Data set isolation for bibliometric online analysis of reserach publications: Fundamental methodological issues, *Journal of the American Society for Information Science*, 48: 205–217.

JAKOBIAK, F. (1985), Bibliometrics and terminological analysis. Some examples using the "Get" command. In: *INFODIAL-VIDEOTEX; Proceedings of the 4th International Congress and Exhibition on Data Bases and Videotex*. Paris, France: Infodial-Videotex. pp. 18–21.

JARNEVING, B. (1999), *The Cognitive Structure of Scientometrics; An Author Cocitation Analysis*, Master's Thesis, Bibliotekshögskolan/Biblioteks- och informationsvetenskap. Högskolan i Borås. Borås, Sweden.

LATOUR, B. (1987), *Science in Action: How to Follow Scientists and Engineers through Society*, Cambridge, MA: Harvard University Press.

LENOIR, T. (1979), Quantitative foundations for the sociology of science: On linking blockmodeling with co-citation analysis, *Social Studies of Science*, 9: 455–480.

LOTKA, A. J. (1926), The frequency distribution of scientific productivity, *Journal of the Washington Academy of Sciences*, 16: 317–323.

LUUKKONEN, T. (1997), Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis, *Scientometrics*, 38: 27–37.

MANNI, P., A. SERRAZANETTI (1987), Online statistical analysis on different hosts: Usefulness in relevance checking and result optimization, *Proceedings, 11th International Online Information Meeting*. Medford, NJ: Learned Information, pp. 367–375.

MOED, H. F. (1991), The use of on-line databases for bibliometric analysis. In: *INFOMETRICS 87/88: Selected Proceedings of the First International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval*, Diepenbeek, Belgium, 1987, pp. 133–146.

MULLINS, N. C., L. L. HARGENS, P. K. HECHT, E. L. KICK (1977), Group structure of co-citation clusters – Comparative study, *American Sociological Review*, 42: 552–562.

MURRAY, S. O., R. C. POOLMAN (1982), Strong ties and scientific literature, *Social Networks*, 4: 225–232.

PERSSON, O. (1986), Online bibliometrics – A research tool for every man, *Scientometrics*, 10: 69–75.

PERSSON, O. (1988), Measuring scientific output by online techniques, In: *Handbook of Quantitative Studies in Science and technology*, A. F. J. VAN RAAN (Ed.), Amsterdam: North-Holland, pp. 229–252.

PERSSON, O. (2000), Bibexcel, http://www.umu.se/inforsk/Bibexcel/index.html

PETERS, H. F. P., R. R. BRAAM, A. F. J. VAN RAAN, (1995), Cognitive resemblance and citation relations in chemical engineering publications, *Journal of the American Society for Information Science*, 46: 9–21.

SALTON, G. (1975), *Dynamic Information and Library Processing*, Englewood Cliffs, NJ: Prentice-Hall. See ch. 10.

SCHVANEVELDT, R. W. (Ed.) (1990), *Pathfinder Associative Networks; Studies in Knowledge Organization*, Norwood, NJ: Ablex.

SHEPHERD, M. A., W. J. PHILLIPS (1986), The profile-query relationship, *Journal of the American Society for Information Science*, 37: 146–152.

SNOW, B. (1993), RANK – A new tool for analyzing search results on DIALOG, *Database*, 16: 111–118.

STEWART, J. A. (1990) *Drifting Continents and Colliding Paradigms; Perspectives on the Geoscience Revolution*, Bloomington, IN: Indiana University Press, pp. 216–238.

THORNE, F. C. (1977), Citation Index – Another case of spurious validity, *Journal of Clinical Psychology*, 33: 1157–1161.

VINKLER, P. (1998), Comparative investigation of frequency and strength of motives toward referencing: The reference threshold model, *Scientometrics*, 43: 107–127.

WASSERMAN, S., K. FAUST (1994), *Social Network Analysis; Methods and Applications*, Cambridge, England: Cambridge University Press.

WHITE, H. D. (1981), "Bradfording" search output: How it would help online users, *Online Review*, 5: 47–54.

WHITE, H. D. (1986), Cocited author retrieval, *Information Technology & Libraries*, 5: 93–99.

WHITE, H. D. (1989), Toward automated search strategies, *Proceedings, 13th International Online Information Meeting*, Oxford, England: Learned Information, pp. 33–47.

WHITE, H. D. (1990), Profiles of authors and journals in information science: Some trials of ORBIT's GET command, *Proceedings, 11th National Online Meeting*, Medford, NJ: Learned Information, pp. 453–459.

WHITE, H. D. (1992), Reference books, databases, and the repertoire, In: Howard D. White, Marcia J. Bates, and Patrick Wilson, *For Information Specialists; Interpretations of Reference and Bibliographic Work*, Norwood, NJ: Ablex. pp. 27–78.

WHITE, H. D. (1996), Literature retrieval for interdisciplinary syntheses, *Library Trends*, 45: 239–264.

WHITE, H. D. (2000), Toward ego-centered citation analysis, In: *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, B. CRONIN and H. B. ATKINS, (Eds), Medford, NJ: Information Today, pp. 475–496.

WHITE, H. D. (2001), Authors as citers over time, *Journal of the American Society for Information Science and Technology*, 52: 87–108.

WHITE, H. D., J. BUZYDLOWSKI, XIA LIN (2000), Cocited author maps as interfaces to digital libraries: Designing pathfinder networks in the humanities, *IV2000, Proceedings, IEEE International Conference on Information Visualization*, Los Alamitos, CA: IEEE Computer Society, pp. 25–30.

WHITE, H. D., K. W. MCCAIN (1989), Bibliometrics, *Annual Review of Information Science and Technology*, 24, Amsterdam, The Netherlands: Elsevier, pp. 119–186.

WHITE, H. D., K. W. MCCAIN (1997), Visualization of literatures, *Annual Review of Information Science and Technology*, 32, Amsterdam, The Netherlands: Elsevier, pp. 99–168.

WHITE, H. D., K. W. MCCAIN (1998) Visualizing a discipline: An author co-citation analysis of information science, 1972–1995, *Journal of the American Society for Information Science*, 49: 327–355.

WHITE, H. D., K. W. MCCAIN (2000), In memory of Belver C. Griffith, *Journal of the American Society for Information Science*, 10: 959–962.

WORMELL, I. (1998a), Informetrics: An emerging subdiscipline in information science, *Asian Libraries*, 7: 257–268.

WORMELL, I. (1998b), Informetrics: Exploring databases as analytical tools, *Database*, 21: 25–30.

ZOELICK, B. (1987), Selecting an approach to document retrieval. In: *Optical Publishing; A Practical Approach to Developing CD ROM Applications, CD ROM, v. 2*, S. ROPIEQUET, (Ed.), Redmond, WA: Microsoft Press, pp. 63–82.

Received March 6, 2001.

Address for correspondence:
HOWARD D. WHITE
College of Information Science and Technology
Drexel University
3141 Chestnut Street
Philadelphia, PA 19104, USA
E-mail: whitehd@drexel.edu