Oral History Interview with Stephen Robertson (SR)

Conducted at Dublin, Ireland, July 2013 (during the SIGIR conference)

Interviewer: Lai Ma (LM)

Transcriber: Annie Hull

Edited by Stephen Robertson

Length: 1:58:00

LM: Where did you grow up, how did you grow up, and what's your family and your early education? I saw on your website that you were born April 6th 1946.

SR: Right

LM: Ok, so can you tell me …

SR: Which makes me one of the so called post-war Baby Boom in UK – as soldiers came back, or service people came back from serving abroad at the end of the Second World War in 1945 – nine months after that there was a baby boom [laughter] as you might expect, so I'm one of them.

LM: OK

SR: I have five siblings – I'm one of six children – one older sister and four younger brothers and sisters. I was born in my grandmother's house in the country, Suffolk, which is on the east coast of England. Although we lived in London I spent all my holidays, for the first ten years at least, and then a bit longer, in my grandmother's house in Suffolk – in the country, on a river. So, what else do you want to know about that?

LM: So you were living with your grandmother…

SR: No, I was living in London…

LM: Living in London…

SR: … going to school in London, but every school holiday for the whole holidays more or less…

LM: So you would go back to Suffolk…

SR: We would go back to Suffolk, yes.

LM: Ok

SR: I had a very happy childhood. On the river, I learned to sail. When I was about ten my grandmother moved – not very far – but to a small town along the coast. That was also on a river, so I did a lot of sailing then, and other things that children do. But I was living and going to school in London, and one of the

schools I went to, from when I was six to when I was twelve or thirteen, was a French school in London.

LM: A French school in London?

SR: My parents thought it would be a good idea. There's a large French lycée in the centre of London, which was originally there, probably, to serve the children of families working in the diplomatic service – not just French but any nationality, a sort of relic of the time when French was the international language, or the language of diplomacy, and so there is this big French school. But there was quite a contingent of English people there, because it was thought to be a good thing to learn French at a young age. Then at thirteen I went to an English public school – what's called public school in England, which is a private school actually.

LM: Okay

SR: [laughs] It's a funny kind of terminology, but it's a rather expensive private school. So I went there from thirteen to seventeen or eighteen, and then went to Cambridge, that's my university, and did a maths degree there.

LM: Yes. So, before we move on to Cambridge, were there any influential people who might have influenced you to take mathematics – or in the future computer science – as a child?

SR: My father was an academic but in a completely different subject. He was a classical archaeologist – Greek art and archaeology was his main interest – but he was an academic, which probably had some influence on what I did. My mother had been a mathematician – although she was a full-time mother. She had done a maths degree originally and I certainly felt more at home in mathematics than in classics or art or anything like that at school, certainly in secondary school. I think I just find maths easy. There could hardly have been anything that influenced me to go into computer science, because computer science didn't exist then [laughs].

LM: It was in the sixties

SR: Yes. I was at my secondary school from '60 to '64, the public school, so computers existed just about, computer science as a discipline did not exist, that's more recent than that. I didn't know anything about computers. I guess in the last year of my maths degree, there was a little bit of computing. And then, immediately after my maths degree I went on to do information science and there was a little bit of computing in that but not very much.

LM: Oh, so that was '67

SR: '67 to '[6]8, yes

LM: '[6]7 to '[6]8 and there was not a lot of computing…

SR: No

LM: So how - what is information retrieval or information science…

SR: Jason Farradane, who ran that department at City University, had been an information officer in a technical industry, Tate and Lyle, which is a big sugar company. The more technical industries, engineering, sciences, medicine, that kind of thing, those big companies had very substantial technical information departments which would have publications – published papers and reports and internal reports, indexed in some form – not on a computer, but searchable through indexes. So the indexes might be some kind of card catalogue, traditional library card catalogue, or possibly some kind of – I don't know if you've ever come across them – punched cards?

LM: [Simultaneously] Punched cards.

SR: There were two kinds of punched cards, one of them is the old Hollerith cards. These were invented for analysing the US census in the 1890s and were mechanically sortable, you could take a deck of cards, and put it into a hopper…

LM: And that was IBM?

SR: That was IBM. There were several punched card companies but IBM was the biggest one, and IBM actually was descended from the original company formed by Hollerith. There were IBM punched cards and other similar things. And then there was another kind of punched cards called peek-a-boo cards, have you ever come across them?

LM: I don't think so, no.

SR: Not for mechanical sorting, but they're rather large cards, with a grid on them, and each card represents an index term – a word, or a subject heading…

LM: Subject headings?

SR: But assigned in such ways that they can be combined, with Boolean AND. So you could say, I want all the documents that contain this term *and* this term, okay? This is how it works, there's a grid, and the grid represents document number IDs, so any number is read off the grid as a sort of *xy* coordinate, for a position on the grid. If you want to index this document represented by this number with this heading, this index term, then you punch a hole there, okay? Then when you want to do a search, you pick up a single card and look at the holes, but also you can pick up two cards and put them on top of each other and look at the holes, and a hole through two cards is a document that's indexed by both those terms, so it's a Boolean AND, okay?

LM: Okay

SR: Those systems – that idea, was actually invented in 1915, for a different purpose, but it was used in a number of technical information departments in the '50s and '60s.

LM: It was invented in the UK?

SR: It was invented in different places actually, the original invention in 1915 was to identify species of birds. There was a US patent. But it was reinvented at least one other time, maybe more than once, for indexing documents in a technical library, one of those was a UK invention. It was used in different places; I know there were American ones as well.

LM: Ok, so it was not encryption that attracted you to information science?

SR: Oh no. I wasn't particularly interested in encryption. I did the maths degree not really knowing what I wanted to do next. I had an idea that I might teach in school. I got married in the end of the second year of my degree, to a teacher – an art teacher – and that sort of put me off the idea of teaching, I decided I couldn't do it – I'm glad I decided that before I tried it, actually, I don't think I could have done it, not in school. So I was looking for what else to do; I was married and I didn't want to continue on the student path, I could maybe have gone on to do research in maths, but I didn't want to do that because I wanted a job. I looked at the work that technical information officers did and thought I might be interested in that and took a one year master's course at City University.

LM: Was it a taught programme?

SR: Yes, it was taught with a three month dissertation at the end – four month maybe. The taught part was very much oriented to people who were going to do what Farradane had done, that is become information officers in industries of various kinds, maybe in some type of library. I wasn't really inspired by that but a chunk of the course was about information retrieval, and I did my dissertation on metrics – evaluation metrics. '68, when I did the dissertation, was two years after the publication of the second Cranfield project, have you heard of that?

LM: Yes.

SR: The second Cranfield project was the one that established the notion that you have a collection of documents, some queries, some kind of relevance judgements and some evaluation procedure for different systems of retrieval. Incidentally the particular systems that were evaluated in the Cranfield project were all library-type systems, so there was the Universal Decimal Classification; a form of faceted classification; a uniterm system – which was a sort of post-co-ordinate system suitable for those peekaboo cards that I mentioned – and something else, I can't remember. All systems that were run, operated, designed by library people. The Cranfield project was done in a rather mechanistic way,

but entirely without the help of machines of any kind, so they had clerks doing mechanistic tasks, in a way which seems very strange now, because they were obvious things to delegate to a computer. And there were computers around between 1962 and 1966 when the Cranfield project happened. But it was run by a librarian, who didn't know about such things, and did know about organising card indexes and getting people to search them in systematic ways, so that was how it was done – completely without machines. By the time I did the information science course, there were some computer based systems, one or two. Do you know about MEDLARS? Or MEDLINE or PubMed?

LM: MEDLINE

SR: Ok, so what was at the time MEDLARS, Medical Literature Analysis and Retrieval Service, was running on mainframe computers in 1966 or '7 maybe, and morphed into MEDLINE, which morphed into PubMed, so it exists in some form. But at the time, in the late sixties, if you want to search MEDLARS, what you do is like this. Suppose I'm a medical researcher in Dublin and I want to search MEDLARS, I write a letter, put it in snail mail, send it off to the National Library of Medicine in Maryland, and then at the other end, a specialist searcher – a specialist intermediary searcher – would formulate my question as a search, as a Boolean search statement of some kind. It would be run overnight on the machines there, and the printout would be sent to me by mail. So, it would be a question of physically crossing the Atlantic twice.

So that existed, and we learned a bit about that in the course, and also the big scientific abstracts journals. Index Medicus which was the basis for MEDLARS is one of them, but publications like Chemical Abstracts and Physics Abstracts were beginning to computerise their production processes – not with a view to searching, but with a view to helping in the production of the printed indexes. But the result of that was the data existed in machine-readable form, and if you were one of the big technical companies running a technical library, you could maybe afford to buy a tape of Chemical Abstracts. That was happening just about in the late sixties and the early seventies. In the early seventies it started happening in a big way, and then one or two companies like Lockheed Dialog started accumulating the tapes from different abstracting services and offering search services to anybody for a payment, which eventually became online search. Not through the Internet, because the Internet didn't exist, but you could have a dial-up connection. Since the computers sat in California, you'd have to have a dial-up connection from wherever you are to California, which was rather expensive, and a slow printing terminal, one of those early IBM terminals perhaps? Twelve character a second is slower than you like to read, so [imitates machine, makes clunking sound] character by character. The first one I used was also a printing terminal, with thirty characters a second, which was about as fast as you could reasonably read, I think probably about '70, '71, or something.

LM: Okay. So you have encountered plenty of experience in all these machines, through your master's programme?

SR: Well, in the years after the MSc. But information retrieval was an intellectual discipline, okay?

LM: Right

SR: Computers might or might not be useful to help you do it but it was about the intellectual discipline. It was about the ideas – most of the discussion at the time about information retrieval was about whether you could represent different subjects with, let's say, a faceted classification, or a hierarchical decimal classification.

LM: Those were the days you were using very discrete systems?

SR: Absolutely, and generating fierce arguments between the proponents of different systems, which was actually what the original Cranfield programme was designed to answer. So, the people who believed in faceted classification would say you can't do it with a hierarchical system, it doesn't work, you need these facets – and vice versa. Cranfield was designed to resolve those arguments – of course it didn't, but it was designed to do so.

LM: Ok. Very nice. So, and then, let's see-you got your master's in '68?

SR: Yeah

LM: And then, you went to ASLIB before you actually did a PhD?

SR: Yes, I did, yes. Do you know about ASLIB at all?

LM: I know Blaise Cronin…

SR: Of course, he worked there after I did – that's right, yeah**.**

LM: You two didn't cross paths?

SR: We didn't cross paths there, no. Aslib was a part public funded, part industry funded, research association for information and we were in the research department, both Blaise and I were in the research department. Which was an interesting place. We mostly did relatively short-term projects, specifically funded projects.

For example I did a project on how facsimile transmission might be used in information services. This was early days of facsimile transmission – there were one or two fax devices on the market. And there were one or two what you might describe as distributed information services: you send off or telephone somewhere remotely – to a technical library let's say – with a request: I want you to find out about this for me. One response might be for them to fax you back some document or other. Except that fax at that time was only suitable for

a single page or very small number of pages. It took quite a long time – it involved printing out slowly at the other end. There were some people who had ideas which now seem quite strange, but in a way pre-date or pre-figure the notion of storing documents digitally. They didn't have that idea but there was a National Lending Library service based in Yorkshire at the centre of England, which was serving a lot of people by post, often photocopying articles and flying them by post on request. The director of that library had the idea that maybe at some point in the future it would be feasible to use fax instead of post, which is a bit like serving it from an electronic database. You'd have to scan it, but maybe you could store the results of the scan. So that idea was around, and I did a report, concluding that I didn't really think that fax was a very good mechanism for doing this, but it was interesting to do the report. So that kind of thing – that was about a three month study.

The then director of the research department was Brian Vickery. He'd come from exactly that environment, he'd worked at the National Lending Library, but he was interested in lots of things about information retrieval and published a number of books which were at the time standard textbooks in the field. He was quite happy for me to spend some of my time doing my own research, and I'd been somewhat fired up by doing the dissertation project – on evaluation metrics for IR.

I got in touch with Karen Spärck Jones at that time – she was in Cambridge doing stuff with the Cranfield collection on term clustering, things like that. I got quite stimulated to do some of my own work. I published a couple of. Of the people in the field from that time, apart from Karen, there was also Bertie Brookes. He was at University College, he'd written something that I'd used in my dissertation project – which I disagreed with actually, I was quite opinionated. But Farradane, who ran the City department, was impressed with my dissertation, and sent it to Bertie. And Bertie said it should be published, so I wrote two papers based on my dissertation, and published them in the *Journal of Documentation*, which was published by ASLIB (Bertie was on the editorial board). Although it took me a couple of years to get round to it, it followed from that, made sense, that I would start a PhD at University College with Bertie as my supervisor, so I did. I probably started in about 1970 or '71.

LM: So you met Karen Spärck Jones even before you went into the PhD programme?

SR: Yes, that's right, I did. I was thinking of other ways that you could analyse the Cranfield results and had some interaction with her on that. I started the PhD part time, I think it was 1970, and still working at ASLIB. In UK, it's hard to do a PhD and get paid a salary at the same time; there isn't the system that there is in the US. There isn't the system of salaried or at least reasonably paid teaching support staff. That notion isn't really there in UK universities, it's

coming in a little bit, but it isn't really there. If you wanted to do a PhD full time you might be able to get a grant to do it but it wouldn't be anything like up to the level of a salary. I wanted to continue working and getting a salary, so I continued working at ASLIB and doing my PhD part time.

There was at the time a scheme for research fellowships in information science. They were called Scientific Information Research Fellowships, and they were run by the Royal Society in London (which runs a number of research fellowships and research professorships and such like). At the time they had two research fellowships, of which the two first holders were Karen Spärck Jones and Nick Jardine. They were very good, very nice research fellowships, they lasted for a maximum of five years, paid a good university salary – in UK terms, a good university salary – equivalent to a junior lecturer, but quite decent, and allowed you to do whatever you wanted to do, more or less. Karen had that fellowship from '68 to '73, and Nick had the same time frame. So in '73 there were two fellowships going vacant. And there were presumably a number of applicants, but I got one and held it from '73 to '78. The person who got the second one was Keith van Rjisbergen –he'd been a student of Nick's but he'd gone back to Australia. He'd postponed for a year, and started the year after me. But so, from '73 to '78 I held the fellowship at University College – you could hold it where you liked – so I went to University College then.

What was a little surprising was that they gave it to me despite the fact that I hadn't yet got my PhD. It was sort of a post-doc fellowship – I mean, Karen already had a doctorate, Nick already had a doctorate, Keith already had a doctorate – but I didn't. But they decided I was well on the way, that I could have it – so I finished my PhD as part of that fellowship, which was very nice. I also continued to discuss things a lot with Karen. She was in Cambridge, I was in London and we didn't have email then, so we wrote letters [laughs]…

LM: Oh, very nice.

SR: …. some of which, I still have. The other person who was at University College at the time was Nick Belkin. Do you know Nick?

LM: Yeah, Rutgers.

SR: Yeah, Rutgers. He was also a PhD student of Bertie Brookes, who started a little after me. When he finished his doctorate, which was a little after I moved to University College, he went to teach at City University where I had been, so he was around. In '75, when he was still at University College, we – that is chiefly Brian Vickery, Bertie Brookes, Nick Belkin and myself (actually Nick Belkin did more work than anybody else) – we organised a research colloquium at University College which involved a number of people, some of whom you've probably heard of, and Karen in particular. Shortly before that, Karen had sent

me, as part of general discussions of things, a paper by one of Gerry Salton's students. And I said, I think I can do better than that.

LM: Ok, so back then there was two different fields. Salton, he was more recognised as a computer scientist.

SR: As a computer scientist. So, there's quite an interesting mixture. Karen worked in a computer lab in Cambridge but had an arts degree and didn't regard herself exactly as a computer scientist. She started working on semantics. Fairly early in the process she got married to Roger Needham, who was a computer scientist, so the two of them worked together, and Karen got a position in the Computer Laboratory at Cambridge (which was then called the Mathematical Laboratory) as a researcher, partly because she and Roger had developed some programs to do clustering and she'd evaluated them. So Karen was sort of between the two. She'd never had the library training, but she did have a linguistics and semantics background, which was then right outside computer science. But she sort of allowed herself to be absorbed into the Computer Lab, alhough I don't think she ever really regarded herself as a computer scientist. And then there were people, for example at the Sheffield University Library School at the time there was a man called Michael Lynch. And more recently, indeed he's still there, Peter Willett, who had been a student of Michael's, doing stuff on…

LM: Systems

SR: Yeah, on systems, including some very technical systems. I'm not sure whether Peter would regard himself as a computer scientist, I don't know actually – he's from the library school, but he's done some quite heavy computing work, including all the stuff on chemical informatics, which is quite a demanding computationally. On the other hand, there were people like Keith van Rijsbergen, who was definitely a computer scientist. His two students at the time were Bruce Croft and David Harper. They all regarded themselves as computer scientists, and in the US indeed there was Gerry Salton who was definitely a computer scientist.

LM: So this was how the overlapping…

SR: That's right, but I think the Syracuse people, who were quite active a bit later, were in the library school, so there was quite a mixture, and it was sort of understood that information retrieval was a problem which spanned those different disciplines so it was quite important for both kinds of people to be involved. And of course, Nick's in a library school as well

In 1975 – maybe the end of '74 or the beginning of '75 – Karen sent me the paper by one of Gerry's students (Karen and Gerry were in correspondence a lot) and I said I think I can do better – there was a piece of theory there…

LM: Is this the vector space?

SR: Well, it was within the framework of the vector space model, but it was about relevance feedback. The idea of relevance feedback had been invented by Joe Rocchio, in Gerry Salton's department. He was certainly the first person to write about the idea systematically. So, this was another of Gerry's students formulating a theory of how to do relevance feedback, and I thought it could be done better. There were a couple of people, both in the UK, whose work I was aware of, which was a little bit similar. One of them was working on how to improve queries, profiles for MEDLARS. And the other one was working with I think the Chemical Information service, one of those big scientific abstracts services, again how you improve queries. If you're running the same query regularly, you see how well it's working, and how you can improve it. With those ideas, and this paper of Gerry Salton's student, I thought I could do something. I put together an idea about term weighting, which I sent to Karen. It was quite a long time, because as I said she didn't really regard herself as a programmer, she wrote and ran programs, but she wasn't an expert programmer – she ran evaluations, she ran ideas for systems, against the Cranfield data. So she tried out my idea, and I got this letter from her, which I still have, saying "it doesn't work at all, hopeless" – and I also have the letter two days later, which said "there was a bug in my program, it actually works very well indeed." And that was the basis for my 1976 paper with Karen, which was what got me into probabilistic modelling and relevance weighting and all that. So, that was published in '76 in JASIST – or JASIS…

LM: JASIS. It didn't have the 't'.

SR: No, it didn't have the 't', I was just trying to remember, it had changed from *American Documentation* I think…

LM: Right, yeah, it was '68 that they changed it from *American Documentation*…

SR: So, it was JASIS, but not JASIST, yes. I have mentioned the research colloquium that we organised at University College – Karen and I spent quite a lot of time during that colloquium discussing this paper, because she'd already got good results, but we wanted to firm it up a bit. So we wrote the paper at the end of '75, submitted it to JASIS, got quite a negative review back, by Gerry Salton.

LM: Did you know that?

SR: Yes, I knew that. That's an interesting change. I can't now remember whether it was officially blind, it probably was officially blind (blind reviewing, not double blind). It probably was, but almost certainly because Karen's name was on the paper, and Gerry and Karen knew each other very well, it became very rapidly a discussion between Gerry and Karen. Gerry did not like it, he

really didn't, he had a lot of objections to it, and there was a bit of to-ing and fro-ing, and eventually he was reluctantly persuaded to let it be published, and it was published, and it's both mine and Karen's most highly cited paper, by quite a long way. Which is nice.

In September '76, that was after it had been published, I made my first visit to the States, did a sort of tour of various places in the States giving seminars – part of it with Nick Belkin. The latter part, when I was no longer with Nick, I did a seminar at Cornell, to Gerry's seminar group. The interesting thing was I had no difficulty persuading his students that it was a good thing – that what I was doing was worth doing. But I still had difficulty with Gerry, he really didn't like it at all, but his students picked it up very easily, were happy with it, so that was quite interesting.

LM: So, all this time you were…

SR: I was still on my Royal Society Scientific Information Research Fellowship at University College.

LM: And you haven't gotten your PhD?

SR: I got my PhD in…

LM: So that was '76, so around the same time?

SR: End of '75, beginning of '76, I did have my PhD at that point, yes.

LM: It was not too much of a milestone because the project actually was more important to you?

SR: Exactly. By the time I got my PhD I was well into the project with Karen. The project appears in my PhD thesis as a sort of afterthought; most of my PhD thesis was still about evaluation. It sort of morphed a little bit into this probabilistic model, which fitted with some of my ideas about evaluation as well, but no, by the time I finished my PhD I was well into doing other things, I didn't worry about it, it wasn't a big milestone.

–

SR: There's something I've missed out, which I must tell you about…

LM: Okay, this is still on

SR: Well, I need to describe one of the features of the information retrieval scene in UK. I should say that UK was one of the major centres for information retrieval – there was some in the States, quite a lot in the UK, some in France, but very little anywhere else in the world, at all. Anyway, the information scene in the UK had the Cranfield project going on – Cleverdon who ran the Cranfield project was regarded as a sort of world authority on this kind of thing – he got

involved in the MEDLARS evaluation in '67 or so, over in the States as well. He was a librarian at Cranfield, what is now Cranfield University, but was then the Cranfield College of Aeronautics, a very specialist college. He had this very specialised library collection, and that's what he used for the Cranfield project. But, he was quite a forceful character, he had very strong ideas, on various things. Jason Farradane, who was the leader of the City course, was also a very strong minded character, with strong ideas of his own. For some reason these two, Farradane and Cleverdon, did not get on at all. Not at all. They almost hated each other, not quite hate, but professionally they they each regarded what the other said as completely bad and hopeless, and at public meetings they'd get up and say so. If one of them was giving a talk, the other would be sitting in the audience just itching to jump up and tell him why he was wrong. There was quite a lot of this.

I became aware of this a little bit when I was a master's student, doing my dissertation. And of course my dissertation was on metrics for evaluation, on which Cleverdon had very strong views. So I did this dissertation, and then published two papers at the beginning of '69 in the *Journal of Documentation*, based on this dissertation. In '69 I was working at ASLIB, but I had been a City University student, a student of Jason Farradane, and this dissertation came out of that. And Cyril Cleverdon got his copy of the *Journal of Documentation* – which he always read thoroughly – as soon it was published, in March of '69, and immediately got on the phone to the director of Aslib – remember ASLIB published the *Journal of Documentation*, and I was working there – got on to the director of ASLIB threatening to sue him for libel for what I said about Cranfield and about him. It was a total emotional reaction, it didn't bear much relation to what I'd actually said, but it was a complete overreaction. So the director – the then-director of ASLIB – managed to calm him down. Eventually as a result of that, in the second part of the dissertation, the second paper which was published June, there are some footnotes on the first page of some corrections to the first part – and those are entirely because Cyril Cleverdon got so upset about this.

So, in this relationship, the fact that I was a student of Farradane's sort of tarred me with the same brush, which was occasionally a bit awkward. Cyril Cleverdon ran these conferences, which prefigure SIGIR actually, he ran two or three of them. And I would go to them, I wasn't around for the first, but I went to the second or third. So I had this sort of slightly uneasy relationship with Cleverdon, who was running them – I'll tell you the sequel to that later on, there's a sequel.

LM: Okay. I'll have to remember that.

SR: Okay, so where are we?

LM: So, we were at a point where you published your most highly cited paper, with Karen and then you got a PhD, from University College, so we were there. And, what happens next, so the paper was very well received?

SR: Yes, yes it was.

LM: But Gerry Salton didn't like it…

SR: Didn't like it.

LM: And then you joined City?

SR: That's right, yes, at the end of my fellowship which was '78. I joined City as a lecturer. Nick Belkin was still there. The department was run by a man called Bob Bottle, who was not an information retrieval person.

LM: That was a natural development in your career?

SR: Yes, that was an entirely natural development. One thing I didn't mention, on the same lines as we were discussing library science and information science and computer science, the department of University College that I was in was actually a traditional library science department.

LM: Okay, so I saw the name as Library Archives?

SR: It was called the School of Library Archive and Information Studies, it had a mainstream – the mainstream was traditional librarianship, people going into national or public or university libraries. But it had a very strong archives section, which is still there. It had a rather small information studies section, which consisted of Bertie Brookes and no-one else. Well, Bertie Brookes and a number of sort of masters' and PhD students of his, chiefly…

LM: He was a mathematician, statistician?

SR: He was a statistician, yes. Just as a side remark – he got into that environment because he had been in the electrical engineering department in University College, but he got involved in a lot of work on teaching technical writing to engineers. Technical writing and other related subjects like search, information retrieval, to engineers, which had more-or-less replaced his main statistical background. He brought some of that work into information retrieval as well, but transferred from the engineering department to the library school because that was the main focus of his work. He published some work on bibliometrics, that was the main application of his statistical stuff. He also published one or two things on information retrieval.

He had me and Nick Belkin as PhD students, and a number of other masters' students doing a variety of things. I was there rather than in a computer science department. And that was a traditional library school. I might have continued there. More or less at the time I finished my fellowship, Bertie Brookes retired,

and there was a possibility that I would take over his role there, which didn't work out. The next most obvious for me to go was City, they were quite happy to have me – so yeah, that was natural. And I stayed at City for twenty years.

LM: That's quite a long time, and you became the head there.

SR: I eventually became head there, yes. There was actually a slightly unpleasant period initially – more about that later. Originally (when it was first formed, when I did the MSc, and when I joined the faculty), it was not a department, it was called Centre for Information Science. During one of many reorganisations of City University, it decided that it should have the status of a department.

LM: So from the very beginning it was information science…

SR: Yeah, it was. Okay, so a bit more history. Until a number of years ago, the professional library information field in the UK consisted of the Library Association, which was a very large organisation representing mainly public, academic, national libraries, with a small but not very active special libraries group. People like Jason Farradane – a whole group of them – decided that the Library Association wasn't for them, it wasn't representing their interests well enough. So they formed a new institute, in '58 I think it was, called the Institute of Information Scientists, which was for that kind of person. They wanted to distinguish themselves from the library profession. Although many of them worked in technical libraries they wanted their profession to be regarded as something different from librarianship, so they adopted this term, information scientist. And regarded themselves as different from librarians, there was a bit of a rivalry there.

City University Centre for Information Science was the one place – the only academic department in the country – which allied itself to that movement, because it was started by Farradane, who also started the Institute. There were at the time maybe fifteen or more library schools around the UK, there are fewer now I think. They all pre-existed, and some of them, particularly Sheffield, took on the ideas of specialist information work. Many of the others didn't, but remained traditional, but they all remained aligned to the Library Association rather than the Institute. There was some mixing up a bit later, rather slowly, but the Institute was always a rather small organisation and eventually was forced – because there wasn't enough support for it – to merge with the Library Association, which became the Chartered Institute for Library and Information Professionals, which is what it is now, in the UK. CILIP now includes the old Library Association and the Institute, but it's still the case that it's dominated by the old Library Association which was always much bigger. I was at a retirement party for one of my old City colleagues last month where several of these former Institute members who were in CILIP, were bemoaning the fact that CILIP

doesn't represent their interests very well, so there's still a certain amount of "aggro" about it.

LM: Okay. So how was the work going on when you first joined at City as a lecturer and then a reader, and then you became the head, so how was that trajectory?

SR: It was a bit of a shock having to teach! I mean, the research fellowship was wonderful, it just allowed me to do research. But I actually enjoyed the teaching a lot. The City department only ever taught postgraduate, that is masters' and PhDs. Which meant that most of our courses, the full time versions, were one year full time – one calendar year, so going over the summer for the dissertation. Which was actually a bit heavy, because we never got full academic holidays, we were always supervising…

LM: Oh, because over the summer you still have to…

SR: Six or eight or ten master's students over the summer.

LM: Theses

SR: Theses, yes. But it was good, I mean I enjoyed that. We – Nick and I mainly – decided that we ought to get a computer in the department, okay? [laughs] I mean, there was a university computer service, but we thought we ought to have one in the department. This was pre-PC, so getting a computer was a major undertaking, but we decided we ought to have one. We got a mini computer (as it would have been called at the time) in probably '80. We started doing a number of computer-based projects. One of the main information retrieval activities at the time took the form of big commercial services like Dialog. Dialog had tapes from fifty abstracts services, and a search system, a Boolean search system…

LM: … a keyword system…

SR: Yeah, that's right. So, a Boolean search system, which required quite a lot of knowledge, understanding, training, skill, to do searches. And in fact, that became one of the main things we taught our students, because our students would become search intermediaries, and search intermediaries were necessary then. So we did quite a lot of training students in online searching. At the beginning we had one terminal, and a very expensive dial-up line to California. That gradually changed over a number of years, and getting our computer helped a bit in that. We had to connect our computer to JANET – the internet didn't exist, but there was a UK system called JANET, which was the Joint Academic Network, which, as far as the UK is concerned is the predecessor of the internet. We got hooked up to that in a rather underhand way. Universities and university computer science departments, or computing services, many of them, not all of them, got connected quite quickly, through official channels, but we didn't have

that. In the early days of JANET, you could only get connected to JANET if you were receiving funding from one of the research councils – one of the government research councils – so it was part of that process. We were never in the running of that kind of funding, because we didn't do that kind of project. But we managed to persuade someone to give us sort of a bootleg line into JANET – which caused some problems later – but we did have one, anyway.

We started thinking about doing front-end experiments, which means we'd have a local program to which the user would talk,which would formulate searches for Dialog or whichever of the big services we were using. We were thinking about doing that, using the Robertson/Sparck Jones relevance weighting, so constructing a query as a set of terms and weights and then generating appropriate Boolean combinations, getting results back and putting them in rank order. Which is a little tricky, but it's possible. And we were doing that in the very early eighties, '81.

I got some money from British Library Research and Development Department, which was a source of funds for information retrieval research, not really big, but there was some. We got some money from that source to run the project, employed a programmer, and then did some user studies where we provided services through this front-end system – actually for medics searching on MEDLINE. Still with intermediaries, we were trying to evaluate what we could do with this system that would make search processes simpler. It was in the early eighties we did that project. That was interesting, got reasonable results, not wonderful. I think we probably reported that project in '84 or '85.

Around the same time there was beginning to be quite a lot of interest in online catalogues – library catalogues – OPACS, Online Public Access Catalogues. This was not generally seen by the librarians as an information retrieval task, it was more seen as a providing access to what had been the card catalogue. The kind of task which librarians were used to using the card catalogue for was identifying a known item, okay, so…

LM: Accession cards and author and title and…

SR: That kind of thing. So, in the early to mid eighties libraries, the big libraries, were just beginning to offer online access to the catalogues in a rather simpleminded way. And Stephen Walker and another small group – not at City, at what was then called the Polytechnic of Central London – were designing an online catalogue system. Steve picked up on Robertson/Sparck Jones weighting and decided it would be a good idea to try to implement something like that in the online catalogue. Because when people started searching they would search for subjects as well as known items. So, the system Steve designed is called Okapi, originally intended just as an online public access catalogue, and it had – which was not totally unique but quite novel at the time – it had a text box into

which you could type anything. I'm almost sure it was the first system in the UK to do that. I'm not sure, but I believe there was another experimental system in the States that did that earlier – provide the user with a text box in which they could type any natural language query. Maybe another front-end system, there were a number of people working in that kind of area. But anyway, Steve implemented it in about '83 or '84 – on this Okapi system at Polytechnic of Central London. And that project ran for two or three years and was an interesting project, but then sort of petered out. That is, the then librarian of Polytechnic of Central London had himself got some money to run this, but he wasn't actually terribly interested in it. Steve Walker and I, and other people who worked on the project, had met and discussed various things, and we decided it would be good to get the Okapi project, move it to City. Which we did at the end of the eighties, maybe '88 or '89, it took a little while. In the meantime I had a colleague, Micheline Beaulieu, who had been doing work on the online catalogue in City University. City University library had an online catalogue, so she had been doing user evaluations with that. We were working in that area, and getting the Okapi system in seemed like a good thing to do. One of the things we did when we got it in was to mount not only the library catalogue data, but also a scientific abstracts database.

LM: Ok, so back in the library there is only the collection of physical books, not the abstracts?

SR: That's right, exactly. Now we had a small set of abstracts – it was a section of Physics Abstracts which dealt with computing actually…

LM: Right, indexes

SR: We had indexes and scientific abstracts which we could search the texts of. So we were moving into adapting Okapi to become a general text retrieval system rather than just an OPAC. And running experiments with users, who were generally speaking students in the computer science department – research students, faculty sometimes. By that time the university computer service at City had terminal access quite widely distributed, so most of the PhD students, for example, could get access to a terminal. And we provided them with access to Okapi and therefore the databases on Okapi, on the basis that they would let us watch what they were doing and evaluate them in various ways. And that was very good. So we had a period of running experiments like that, and finding out a lot about user searching and things like that.

To backtrack a little, in the early '80s, I had started within the department a research centre called Centre for Interactive Systems Research, which was to house the front-end project, and then when it came along, Okapi.

LM: So it includes applying weighting to searching on the Dialog system?

SR: Yes, that's right, the front-end project. It didn't have a very strong formal existence, this Centre, but it was a group of people who were working together, people I was able to employ on the grants that I had plus students, plus one or two faculty, including Miche. Nick moved to Rutgers in the middle '80s. By the time we brought in Steve Walker and Okapi, '88 or whenever it was – which was also the time I became Head of Department – that Centre was a home for the Okapi project, a series of funded project based on Okapi. Then in '90 or '91 TREC (the Text REtrieval Conference) was announced. But again let me backtrack a little. I'd actually never myself got involved in running test collections like Cranfield. As I said, Karen did a lot of experimentation and over the seventies there were a number of additional test collections which Karen accumulated, and she would often run experiments on some of my ideas as well as others on these test collections.

LM: So you were working on Okapi?

SR: Okapi was not a test collection, we didn't try to do systematic relevance evaluations. We got some relevance evaluations, but only as part of the interactive search process. We were working a little away from the test collection world. In the late seventies, Karen and Keith van Rjisbergen put together a report on the 'Ideal' Test Collection, 'ideal' in quotes, which was about the fact that all the test collections that we had, Cranfield included, had been designed for specific experiments. Considering the kind of use we were making of the test collections – or that Karen and other people were making – we thought we ought to try and design a test collection that would be more or less general purpose, reusable, and that was what the 'ideal' collection project was about. It was funded and operated almost entirely within the UK. Gerry Salton must have got involved, and one or two other people outside the UK, but it was funded by one of the UK agencies, this study of how best to do a bigger, better universal test collection.

Karen and Keith did a lot of work on an initial report, and then Karen did a lot more work on further analyses. The end result of that was a proposal that said let's build the 'ideal' test collection, at this cost. This cost being – I can't remember what the amount was, it wouldn't make sense in today's terms – the point was that it was more or less the entire basic research budget for information science for the UK by this agency for two or three years. Inevitably it didn't happen, I mean it was too much to take on. So the 'ideal' test collection report was put on the shelf and gathered dust.

Then at the tail end of the eighties, early nineties, Donna Harman and others in the States were discussing the idea of better test collections for text retrieval (the genesis of TREC). Dave Lewis said to them, look, there's this report that's sitting on the shelf there, which tells you how to do it. And Donna and people took that up – I mean they didn't do it exactly the same way by any means, but they took a

lot of ideas from it. But TREC had funding from security agencies, so the funding was there, it was a bigger funding operation altogether.

LM: The best dressed people in the conference would be from the FBI [laughs]

SR: Exactly. And they sat at the back and never made any comments [laughs]. So, FBI or CIA or NSA, I think actually the initial funding was NSA, but we were not told at the time.

LM: That makes sense

SR: [Laughs] Yes, yes. Anyway, so the funding was there, that was the point, and also an extremely good organiser in the form of Donna Harman. So that idea became a reality – TREC. When we saw TREC announced, we said, we're going to take part in that; that looks good. We were dependent on getting grants to do particular projects, particularly if they involved equipment and programming staff. We went to the British Library and said we would like to take part in this American but international competition. We were the first UK people to ask them, and they funded us – actually they funded us for several years. Not on a very large scale, but they did fund us for several years to take part in TREC, and felt exonerated in that by the fact we did so well in TREC-3. So we were lucky, because a few years later there were several UK groups who wanted to take part, and some did, but it was much more difficult for them to get public funds, because the public funds that were available had already been given to us.

Then we had to convert Okapi from being a purely online interactive system, so it could do offline experiments, which was essentially what TREC was about. We did that in'91/2, but we didn't do very well at all in TREC-1. I knew perfectly well why we hadn't done well. We had problems with document term frequency – we didn't have that in the model. The Robertson/Sparck Jones model doesn't use term frequency. Gerry had been doing experiments with term frequency for a long time before that, and it was clearly useful, but I didn't have the model to do it. Around the time of TREC-2 I developed a model which became BM25 – didn't have it quite ready in time for TREC-2, so we didn't do very well – a bit better, but not that well – in TREC-2. In TREC-3 we did very well indeed, better than all the other systems, including Smart and Inquery, and the others. So, that sort of put BM25 on the map. That was a few years before the language models started coming in, they came in '98, TREC-7.

LM: So BM25 is still…it's still quite….

SR: It's still the one that everybody else wants to beat, yes [laughs]

LM: So there were quite a few projects, but very oriented to information retrieval. Perhaps, can you tell me how you became the Head of Department, and how that influenced you?

SR: I said there was initally a slightly uncomfortable stage because when the Centre became a department, the rule about departments was that heads of department were appointed for five years. Bob Bottle who had been permanent Director of the Centre, became Head of Department for five years, and as we approached the end of five years, I was invited to apply. Bob Bottle would have been quite happy to continue, and in a way, thought it was his right to be reappointed, but…

LM: But then you were reluctant to apply for the post?

SR: I was invited to apply by both other people in the department, and also by the Academic Registrar, one of the senior people at the university. Bob Bottle was a slightly awkward person to get along with in many ways, and I think they wanted a change. And I did apply, slightly reluctantly, and I was appointed and Bob Bottle remained on the faculty, but took it very hard. He was very annoyed that I'd stood against him, and that I'd been appointed. I think he thought it was entirely improper. Until he retired a few years later, we had a very awkward relationship; he went off into a corner and did his own thing.

So that was how and why I was appointed, and it was partly because I was a successful researcher, certainly the most successful researcher in the department. It wasn't a very large department, but I was the most successful researcher, bringing in funds for research projects as well. That was the sort of thing that appeals to people, so I was made Head of Department. I mean, in a way, it's contradictory, because as Head of Department you have less time to do research. However, I remained Head of Department, and we had – at the time in the UK – what's now called REF, but was then called RAE, Research Assessment Exercise, which was an evaluation of all UK academic departments. During my eight years as Head of Department, there were two RAE exercises, and we got the best possible rating both times, on the strength of the research of the department as a whole, but it was certainly my group's that was the strongest part.

I remained Head of Department until after Bob Bottle retired, and after these two exercises and then Miche Beaulieu who was on the faculty took over for a short while, so that was '96. Then, in '97, I got an email from Roger Needham, Karen Spärck Jones' husband, saying I'm starting this Microsoft Research lab in Cambridge, would you like to join? At that stage, I'd had quite enough of the negative things in academic administration, which are having to deal with committees at all sorts of levels, having to increase our student numbers every year to keep the budgets balanced, many applications for a small number of small grants, that kind of thing. So, I'd had enough of that, Microsoft research seemed like a place where I could go back to doing research on what I was interested in, without all the hassle of academic administration.

LM: Okay, so that was the reason?

SR: That was the reason, yes. I mean, it was okay the twenty years I spent at City, but after twenty years, it felt a bit much.

LM: So do you think the character of studying will have changed after you take the leadership?

SR: It changed over time quite a lot, although at the time I left, its bread and butter was still mainly master's courses for people who would become technical information officers. Intermediaries, search intermediaries often, but other kinds of information officers. It had broadened out slightly from that. We (a colleague and I) had started a course called Information Systems and Technology which was actually about converting people who didn't have a computer science background to some aspects of computing. Not so they could become computer scientists, but so they could bring computing into whatever else they were doing. On that course we had a lot of further and higher education lecturers, who might have previously trained in any subject under the sun, but saw it as necessary for the development their careers and their subjects to learn more about computing. That was an interesting course too, outside of traditional information science. After I left, obviously courses changed, but I haven't really kept up.

LM: Was there a doctoral programme back then?

SR: Yes. We had doctoral students, not very large numbers of them, a few doctoral students over the entire period. Of whom I had four who are still active in the field. One of them is Jimmy Huang, who is at York University in Canada. He was Chinese, he came to me straight from China, around TREC-6 when we had a Chinese track in TREC, which made quite a good combination. Andy MacFarlane who is still at City, Ayse Goker, who was at City until recently, but has now moved to Robert Gordon University in Aberdeen, and Olga Vechtomova who is in the University of Waterloo, in Canada again. And, a fifth one Efthi Efthimiadis, who died a couple of years ago, having been at the University of Washington. There are a few more who are no longer active the field, so there are a few PhD students of mine. I don't have anything like as many as Keith van Rijsbergen who was turning them out in much larger numbers, but I have a few.

LM: Okay, and then you moved to Microsoft Research until very recently.

SR: Until last month, yes.

LM: Last month [laughs]. So, how is life like now, because I know academia much more than industry research, so how was it? It's a good twenty years too, yes?

SR: No, fifteen. '98 to '13.

LM: Oh that's right, fifteen years.

SR: That research lab was started in '97 just before I joined. Roger Needham was a long term academic computer scientist at Cambridge, head of the department (Computer Laboratory) for quite a long time. And he ran Microsoft Cambridge lab very much as a sort of academic research place. He was interested in the sort of things that academic researchers are interested in, like publication. As a researcher one was expected to write and publish papers, and go to conferences and all the rest of it, in much the same way as academics. And as a by-product, maybe we would provide something that Microsoft could use. That was Roger's attitude, quite strongly, and that was the basis on which I joined. I really liked that. After he died – he died while he was still in post – the subsequent directors – we're on the second since him – moved it more in the direction of helping product groups being one of the main aims.

I formed a small group – of two or three people – for information retrieval research, and one of the things we did quite early on – not straight away, probably early 2000s – was to talk to the people who ran something which was then called Index Server, a Microsoft Product for indexing and searching things on an intranet or company network. It wasn't a very good product. We went along to them and said you need better search algorithms than that, and gave them the BM25, which they used among other things.

A bit later, one of my colleagues in my group at Cambridge, Hugo Zaragoza, went back to look at what the product group was doing and realised they hadn't implemented the BM25 properly; they'd made a mess. And in particular, they'd made a mess because they wanted to cope with different fields of the record. Things like title / abstract / body of text / added keywords, that kind of thing.

LM: So very traditional ways to organise a record.

SR: Exactly. And BM25 as it stood – as it was developed for TREC and as I'd published it – was not aware of fields, and didn't have any ways of dealing with them. So what they'd done was they'd implemented a separate BM25 for each field and then combined them, which actually doesn't work with the BM25 model. So we told them we'll work out a way of doing it so it does work, and we worked out something called BM25F which was field aware BM25, which we published in 2004. That worked much better, so they were much happier with that. At the time Microsoft had various search engines buried in different products, like Outlook or Office for example, and had various plans to use some common search algorithms, but wasn't actually doing that. At the time Microsoft didn't have its own web search engine either.

LM: Bing was not there yet?

SR: Bing was not there yet, no. The MSN website operated a search, that's right, but it was contracted out to a company called Inktomi. Inktomi ran a search engine on Microsoft's behalf. Around 2005, 2004 maybe, Microsoft realised –

rather belatedly, in Microsoft fashion – that web search was actually getting quite important. 1998, when I joined, was the year that Google started. AltaVista and Lycos and Yahoo were already around, but Google started in 1998. Yahoo, I believe, was the first major search engine to implement (in 2001) the sort of click-through model of paying for advertising (although the idea had also been around in some form since 1998). The idea is that you associate advertising results with search, and base payment on click-through, okay?

LM: Alright

SR: But Google picked it up quite quickly after Yahoo, and made a real success of it. So by about 2004 it was obvious, not only that Google was very successful, but it actually had a very successful business model. Which, Microsoft suddenly thought, ah yes we want to be in that space. So they brought the search service in-house, and they made two attempts to brand it as a new thing. Before Bing there was something called Windows Live, with Live Search as the search service. But then it became Bing, and since then it's done reasonably well, better in the States than over here. It's a long way behind Google, in market terms, but it's not that bad. It's got a significant portion of the market, and seems to be reasonably good. Arguably, reasonably good. The first version of the Microsoft web search service, which became Live, was developed entirely in-house. I wasn't involved in that at all, it was developed in Redmond. So, one of the things about the Microsoft lab in Cambridge, is that it's actually rather far from where most product development happens and that encourages the academic side, discourages the product development side.

LM: The main product development is in Seattle.

SR: It's in Redmond outside Seattle, that's right. So, that, the people who first put out the web search service for Microsoft had some ideas about machine learning that were rather crude and some ideas about ranking algorithms, but rather simple-minded. Well not simple, no, they were complex in the sense that they had a lot of different features, but sort of relatively simple-minded ideas about how to combine these features and whether you could make any sort of theoretical model. At some point around 2005, I – my group at least – became sufficiently involved that we said you really ought to have BM25 in there – or BM25F actually, the field version – as one of the features, at least. One of my small Cambridge group wrote BM25F as a program, in a form that could fit into their system, took it over, and they installed it and then started including it as one of many features, and [after training] it became quite rapidly – by quite a long way – the most important feature. So then they adopted it.

What happened after that was that the machine learning group – a very strong machine learning group in Redmond/Seattle – developed a new way of combining features, based entirely on machine learning. This is machine

learning using Cranfield-like evaluation data; that is queries and relevance judgements. Not judged to the same depth as Cranfield or TREC but some of those judgements. Machine learning was used to come up with weighted combinations of features or some way of combining features. The first one was in 2006 combining features, in which again BM25 (actually by this time there were multiple BM25 features) BM25 features were very strong components, but they were combined with other things. There are lots of things that BM25 doesn't address, like it doesn't address adjacency or proximity of words in the text, and it doesn't address anything like page rank or….

LM: Or personalise the search

SR: No, it doesn't, that's right

LM: That's much later

SR: Yes it is, that's right. So there are a number of things that it doesn't address, and the algorithms that Bing uses now are fairly complex, combining a number of things, but BM25 is in there somewhere.

LM: It's still a centrepiece.

SR: So at various stages we advised them. We continued to advise the people running the intranet search facility, which still exists quite separately from Bing. It's now called, for a long time it's been called SharePoint search, it's part of the SharePoint system.

LM: A lot of people, a lot of institutions…

SR: A lot of institutions use SharePoint. As part of SharePoint you get a search system which is based on the one we contributed to some years before.

As the Cambridge lab became more associated with product groups what had been my research group within Cambridge transformed into a small group of applied researchers for Bing, and another small group of applied researchers for SharePoint. So by that time, a couple of years ago maybe, I really didn't have a group except informally: I didn't have any management responsibilities for a group because these applied researchers reported to line managers within the product groups. But they were located in Cambridge because it was thought good for them to interact as a group, and with me. That had been going for some time and worked very well. The group still exists now, and because it's a concentrated group in Cambridge, it's still the case that product groups want to keep applied researchers there. One of the people I had as part of my group was Nick Craswell. He was Australian but he eventually went to Redmond, was persuaded by his boss within Bing to move there. But he keeps very close touch with the Cambridge group and he currently has two of these applied researchers

reporting direct to him. They're working in Cambridge and reporting to him, both of them are here at SIGIR.

LM: Wow. So that's a wonderful career. So with all this time you are still associated with City, and University College, in certain ways?

SR: In certain ways. When I left my full time position at City in '98, I retained for quite a long time a small part time appointment there, so I continued to be paid by City for a while. That was regarded by Microsoft as a good thing for some of their people to do. I retired from that a couple of years ago and became Professor Emeritus, that's an honorary position. So I could go and do stuff there if I want to. At the moment, there's nobody that I'm collaborating with there. I've had an on-off relationship with University College for quite a while, but one of my Microsoft colleagues recently moved to the Computer Science Department at University College. She's someone I've done a lot of work with, so it seems like University College would be a good base for me to continue to do whatever I feel like doing. So I now have an unpaid visiting position at University College. I don't know quite yet how I'll use it, as I said, I only retired last month. A lot has been happening now, and is happening over the summer, so come September I'll work out what the pattern of my life is going to be. I expect it'll involve going into University College and talking to people there, and maybe getting involved with some students there. For three years I've been jointly supervising a PhD student at University College, and that's been very good.

LM: Alright, wonderful. Now, let me check this, and hopefully it's still working [laughs]. Okay, now, I think this is really good, and really, really helpful. Obviously this is not going to cover a lot of things in your life, so do you have anything to add?

SR: I tell you what, I didn't finish that…

LM: …that story

SR: I didn't finish that story. In about '96, we, my group at City, wrote a series of papers which became a special issue of the *Journal of Documentation*, about Okapi and the work on Okapi and the TREC work and the interactive user work and so on, a whole series of things, as an entirety. It was published in'97. A few months afterwards, I got a long handwritten letter from Cyril Cleverdon. He'd retired long since, he was quite old at that point – it seems extraordinary that he was still reading *Journal of Documentation* from cover to cover, but obviously he was. In a slightly begrudging, but actually very generous way in the end, he was saying you've done a lot of interesting work there – this was the guy who'd rung up Aslib and threatened to sue for what I'd said about him. He still had some acerbic remarks to make, for example, I said that TREC was the grandchild of Cranfield, via the 'Ideal' Collection project, and he said "downhill all the way" [laughs]. That's a typical catty Cleverdon remark, but at the end he was really

quite generous about what I'd achieved in the – let's see, it was '96, it would have been twenty seven years or so since he'd rung up Aslib [laughs].

LM: So were you very pleased?

SR: I was pleased yes, it was really nice actually. I appreciated that very much. I don't know what prompted him to write, but anyway he did.

LM: Okay, so one last question, this is more of my personal interest. I read your Salton Award lecture, and you said that you have considered yourself more of a theorist than a practitioner – in a very practitioner field – in many ways. So can you tell me a little bit about how you think about what is a theorist and why you consider yourself as a theorist?

SR: I describe my relationship with Karen Spärck Jones, which was that I was the one who theorised – in respect of the '76 paper anyway and in other areas a little bit. I produced a piece of theory and she implemented it and tested it and extended its use, did a lot with it, but it was originally my theory. Nearly twenty years later, when Steve Walker was at City with the Okapi project and we had started on TREC, I was working on the theory of BM25 and Steve was implementing my ideas and running the experiments. I was very happy, because I'd never got involved with setting up or even running an experimental system like Smart or Inquery or the various other systems to enable IR researchers to run experiments – Lemur and Terrier are other examples. So many researchers in information retrieval get involved in setting up experimental systems, to test their ideas and other people's. And I never did that.

We did have the slightly awkward front-end project, where we built front-end systems but the back-end, the search service, was provided by someone else. That was always a little uncomfortable technically, I mean you were relying on things which you didn't know were stable. So for example, DataStar, which was a service we were using, had a minor change in their syntax which messed up our programs completely until we worked out what was happening. That kind of thing. I've always relied on other people to do experiments and to implement things.

I feel most at home with developing theoretical arguments and theoretical ideas. I've got a paper in a conference called ICTIR in October, ICTIR is the International Conference on Theory in Information Retrieval, which is a spin-off from a SIGIR workshop. It's been running as a biennial conference for the last few years – this is the fourth. It's a relatively new conference – there is a lot of theory at SIGIR, but to get a paper into SIGIR, you probably want some theory, but you also want some experiments to show that your theory is better than anyone else's. ICTIR is a bit more open to theoretical arguments which don't necessarily result in experiments.

My paper at this year's ICTIR actually comes out of my master's dissertation – 1968 dissertation, 1969 papers. There was a model which was proposed as a way of looking at evaluation results by someone called John Swets, originally in '63, there was a '69 paper about it as well. I'd written about it in my master's dissertation. There was a paper by Bertie Brookes in '68, which I also included in my master's dissertation, which said, here's a way to use the Swets model, and I said in my dissertation that I didn't think this was a very good way to use the Swets model because of a particular theoretical problem. Which, despite the fact that I published that in '69, everybody ignored, and for the last decade in particular, there's been quite a lot of work on the same model, applying that model in the way that Bertie Brookes applied it, with exactly the problem that I'd identified in 1968. And I've been a little bit involved in that work in one or two ways, but the problem's been niggling at the back of my mind. I came up with a theoretical solution to it last year, and that's the paper that's appearing in ICTIR.

I would say most people in the field – even the academic researchers – are enthused by trying things out and experiments and actual systems and so on, rather more than they are enthused by theory. It's something that's just been missed, I think, and it's the sort of thing that does interest me. I still don't think that information retrieval is a very strongly theoretical discipline, but there are areas where a kind of theoretical argument can contribute significantly, and I'd like to be able to look at those, and they appeal to me.

LM: Okay, wonderful, yes. And do you think that this is something that's maybe lacking in the current education…

SR: Yes, I think it is. When I started, the students that I taught, and the other students on the course, on the MSc that I took, most of them were going into jobs in practical information work. And nowadays, practical orientation is different, because there are a lot of computer scientists whose practice is to get involved in programming and so on. So what is practical is different now than it was then, but it is the case that most students in the field have one or other kind of practical orientation. If you don't naturally develop an enthusiasm for theory it's hard to teach. So I imagine the kind of orientation I have will be in the minority anyway. I think there'll continue to be some people who are interested in it…

LM: I do hope so

SR: That's right, and I hope they'll continue to make themselves known in one form or another.

END OF INTERVIEW

Postscript

One phase of my professional life that I skipped in the interview happened in 1981. I took a short sabbatical from City, and spent three months in the Library School at the University of California Berkeley. I taught a class there, but my main reason for going there was to collaborate with Bill Maron and Bill Cooper, both working there at the time. Bill Maron had been a co-author of the first probabilistic model for information retrieval, published in 1960, and he and Bill Cooper were working on some related models. Both Bills were quite naturally theory-oriented, and the theoretical challenge we set out to resolve had to do with the relation between the Maron/Kuhns model and the Robertson/Sparck Jones model. It turned out to be quite hard, and while we identified the fundamental theoretical problem, we were only very partially successful in resolving it. We published two papers on our ideas.

Once again, it's a problem I have returned to at intervals. I believe that we have subsequently taken some major steps towards a solution, particularly very recently, but it remains a difficult theoretical problem. It also now has applications, in web search and in many other kinds of systems, way beyond what we imagined in 1981.