Zdzisław Pawlak

1960ies and 1970ies

V.W. Marek

Department of Computer Science
University of Kentucky

September 13, 2016

Zdzisław Pawlak 1926 – 2006

- This session commemorates life and work of Professor Zdzisław Pawlak
- Likely I am one of few active collaborators of *Zdzisław*
- Moreover, I worked with him during the “heroic” period of his activity, when he introduced *Information Storage and Retrieval Systems*, and at the beginning of his work on *Rough Sets*
- But it all started earlier ...

Foundations in Warsaw, 1960ies

- I started at Warsaw University, Mathematics, in 1960
- Very soon I was attracted to *Foundations of Mathematics*
- There were several seminars at Warsaw University that dealt with Foundations:
  - Mostowski's Seminar
  - Rasiowa Seminar (Algebraic Foundations)
  - Szmielew Seminar (Foundations of Geometry)
- And then there was a seminar, led by Bartoszyński, Ehrenfeucht and Pawlak at the Mathematics Institute of the Academy of Sciences

What's so special bout that seminar?

- Entirely different attitude to science
- No hierarchy, no 'Sir, Madam'
- Excursions into unknown
- And the understanding that computers form an entity that eventually will need to be understood and developed
- There were, of course, computers - after all Zdzisław built some (and others followed) - but *understanding* of computers was a different matter

Few examples

- The audience was truly interdisciplinary, with mathematicians, computer scientists, biologists, medical professionals
- The net effect was that the matters discussed at the seminar were going a bit out of a mainstream (as it was then)
- For instance Pawlak discussed the grammar of DNA as a formal language (when I hear today about the quest for creating (as opposed to sequencing) of a valid human genome, I wonder if it was what Pawlak had in mind)
- Question of a possible transmission of biological information through various media was discussed
- (I recall making a presentation on Turing Machines)

Doctors at the seminar

- Several medical doctors (this included, as I remember, Professor Jan Doroszewski, MD) discussed potential applications of computers in Medicine
- This had an important influence on work of Pawlak (and I will discuss it below)
- The issue was if computers would be able to analyze medical information
- Today, it is obvious - we use computers and all sorts of techniques to analyze the medical information. It was not so obvious in the 1960ies
- But before one analyzes medical information, one needs to *describe it* and have it *stored*

How would we store it?

- We need to remember that we talk about 1960ies, pre-SQL, relational model, etc.
- Understanding that data (essentially any data) could be stored was already present in the community
- In fact there were at least two dominating paradigms for storing data called “hierarchical model” and “network model”
- But both required involvement of IT staff to set up storage and retrieval
- If you wanted (say, medical) professionals to handle the data this was out of question
- (With all due respect, an MD designing a double-linked list?)

Some distractions

- But recall that I am a professional logician, and was a graduate student in mathematical logic in the 1960ies
- Several things happened. A mathematician from Stanford, Paul J. Cohen, invented a technique called “forcing” and all of us, students of Professor Mostowski, dropped everything and dealt with this technique (whatever is it ...)

Distractions, cont'd

- Andrzej Ehrenfeucht (most important mathematical advisor of Professor Pawlak) emigrated to US and eventually settled in Boulder, CO.
- The entire period 1967-1970 was not especially conducive to doing research (with the war in Middle East and its consequences - particularly in Poland)
- And so, for a couple of years I was mostly busy with Ph.D. and mundane task of surviving
- Then, in 1968 I got my Ph.D., and in 1970 went as a postdoc to Utrecht, Holland

But not everything was lost ...

- Actually, Andrzej Ehrenfeucht had two students: Grzegorz Rozenberg, and myself
- I went to study with prof. Mostowski. Grzegorz Rozenberg emigrated to Holland and in Utrecht we met again, and - actually - we shared the office
- Unlike myself, Grzegorz was a bona-fide computer scientist (interested then in biological applications) and I heard from him about various computer science problems
- Also, Henk Barendregt was in Utrecht, and so I learned about λ-calculus
- But I was supposed to write a habilitation, and I did

There was life after Utrecht

- As I was coming back from Utrecht, I decided to visit another Polish logician, Dr. Janusz Onyszkiewicz, right then in Aarhus, Denmark
- Much to my surprise I found that logicians in Aarhus were *all* involved in Computer Science
- What happened in the meanwhile was that logicians understood that foundational problems involve computation and computers
- (Not that the top of the profession was not aware of these shifting priorities)

But what was happening in Poland?

- Many things changed in Poland as the result of the so-called “December Revolution” of 1970
- In the area of Computer Science and its foundations a number of steps were taken
- There were several forces and personalities involved
- Certainly, Professor Pawlak and also Professor Rasiowa were at the forefront of the change
- Let me mention various specific events that contributed to that change

Small steps resulting in big changes I

- An organization called “Computational Center of Polish Academy of Sciences” changed its character, becoming, in effect, Computer Science Institute
- The change did not happen overnight, but a group of individuals including Professors Blikle, Dębiński, Mazurkiewicz, Winkowski, and others joined the scientific personnel of that new entity
- (Formal change of the name happened later)
- (Other institutes of Academy also introduced Computer Science themes)

Small steps resulting in big changes II

- Polish Mathematical Society introduced new series of its Proceedings, called *Fundamenta Informaticae*
- (This did not happen overnight, and required a lot of political skills)
- Its name and character were pointing to great traditions of Polish Mathematics, but stressing the need for development of Foundations of Computer Science in Poland
- Both Professors Pawlak and Rasiowa worked over the years (they certainly had help, let me mention Andrzej Skowron and Damian Niwiński) to make *Fundamenta* an important venue for theoretical and practical Computer Science publications

Small steps resulting in big changes III

- Numerous universities introduced faculties that prepared Computer Science cadres
- This included my Alma Mater, Warsaw University, where the Faculty of Mathematics and Physics split, with one of the succeeding faculties renamed (eventually) to Mathematics, Mechanics and *Informatics*
- Several technical universities started faculties devoted to Computer Science (eventually, all technical universities introduced such programs)

Small steps resulting in big changes IV

- But the most interesting and one that had the furthest consequences (at least for me) step occurred at Warsaw Technical University
- Here is what they did: they introduced a program called *Technical Physics and Applied Mathematics*
- Whatever was planned, the result was a world-class program in Computer Science
- That program was immensely competitive, and catered to the “cream-of-the-cream” of the students with mathematical/technical interests
- A group of first-class individuals, winners of science olympiads, from entire Poland joined the program
- Today many of them are leaders of Computer Science - in Poland, but also elsewhere

Dropping some names ...

- The list of alumni of the FTiMS program included: Witold Lipski, Mirosław Truszczyński, Tomasz Traczyk, Zbigniew Lonc, Wojciech Ziarko and others
- The point is that they and (more generally) new generation of Polish computer scientists were ready to bring the area (CS) to the world level
- While most of that initial group left during the years of troubles, the legacy lives
- (Today, Polish national prize for young computer scientists is named after Witold Lipski, the leader of that group)

Back to science

- This story mixes science and personal life, now let us look at science
- I mentioned the medical motivation of the work by Pawlak in late 1960ies, early 1970ies
- The point was to use computer science means to enable research in areas other than CS
- The issue was: “How can we support research of others using computers?”

Storing and retrieving data

- Pawlak proposed the concept of *Information Storage and Retrieval System*, a data structure to organize data so it could be operated by non-CS specialists
- (He was not the only one to think in these terms)
- This was done in parallel to *Relational Model of Data*
- While RM stored data in a collection of tables with relational operations of selects, project, and join (and then many more), the ISR were based on *single* table processing

Storing and retrieving data, cont'd

- One can think about and ISR as an *universal instance*, a concept that was intensly studied by RM researchers
- Universal instance (conjoining all tables of a database) was rejected because the size of the resulting instance was too big
- To deal with this issue RM introduced normalization of tables (various normal forms) to trade computation for storage (which was expensive at the time)
- (Of course today, often we denormalize databases)

Storing and retrieving data, cont'd

- ISR also dealt with the issue of database sizes, but tried to use combinatorics ("consecutive ones property") to handle the issue of instances that were too large.
- There was one more important difference with RM: ISR admitted duplicates, table was a *bag*, not *set* of records
- In other words, the object ID was *not* the part of query language

Motivations

- The issue was the motivation: Pawlak's idea was (certainly a simplification) placing a monitor in doctor's office with medical personnel storing and processing data *themselves*
- Today it is a "normal practice" (at least in US) but think 1970...
- Moreover, while RM, in principle, provides a foundation to modern databases, anyone who *teaches* DB knows that it is only a crude approximation of the reality of modern DBMSes
- This is still an issue in *practical* applications
- (Medical databases and their interfaces are still an issue today)

And so we all looked at ISRs

- For a number of years the notion of ISR was studied very intensly by Pawlak's circle
- As usual in any research venture, papers were written, seminars conducted, implementations tried, dissertations presented, etc.
- The connection with RM were soon noted and investigated (Lipski, and his student, Imieliński)
- We need to recall that all this did not happen in the void – there was still "cold war" around, and world was quite divided
- But for us, Pawlak's younger collaborators, it was a truly international effort

The language issue

- One can think about objects stored in ISR system in linguistic terms
- There is more than one way to do this
- One can think about an attribute as a function assigning to an object a value (for instance: *lname* assigning to objects their ‘last name’)
- This generates (like in SQL) description of an object as a tuple
- And, of course, different objects may have same descriptions - if the language is not very rich
- The choice of the language is left to the user

But this may be misleading

- The point is that language may be not adequate to description of objects
- Medicine is (and better: was) a domain where this problem is acute
- Attributes values may provide a “shallow” description
- (“The patient o1001 has a cough”, but not what is the reason for the cough)
- In other words: the choice of the description language may cover our ignorance of the “real problem”
- But once you are thinking in this terms, it is natural to think that description of objects shows *similarities* between objects

One way of interpreting this phenomenon

- When we try to abstract from the description language we have a natural way of thinking about the ISRs, namely that we have an equivalence relation on the universe
- The equivalence relation is: “within the *query language* used at present the objects have same description”
- The language may change: with a more expressive language the corresponding equivalence relation becomes finer
- It is a well-known fact that the collection of equivalence relations over a given set forms a (non-distributive) lattice
- And, of course, since 1930ies and the work of Birkhoff, *a lot* has been learned about that lattice

The issue of the language

- Often we use some description language (think medicine and symptoms) but we are aware that in reality there is a different language which could be used for description of objects, but for whatever reasons we can not use it
- (In medicine it could be, for instance, description of the case in terms of a biochemical processes)
- That second language may be more precise, but for whatever reasons (cost, availability of testing devices) we use the simpler one
- It should be clear that if the first language is interpretable in the second one, the corresponding notions of ISR are related; the "finer partition" relation can be seen

Rough Sets

- So, by now, it should be clear *how* and *why* Pawlak invented Rough Sets
- (I was present during the discussions leading to the invention of RS, so I know)
- Once you start thinking in terms of equivalence relations (thus partitions) many things become clear

Rough Sets, cont'd

- The point was that once the basic data structure (equivalence relation) was introduced, the connections to many areas became immediate
- These included: statistics, universal algebra (Boolean Algebras with operators), combinatorial optimization, topology, modal logic and other areas
- Actually, new connections are published quite often
- But importantly, a new way of thinking - which is often called *Rough Sets* point of view - was introduced

More on relationships

- Tools such as (for instance) Johnson and Lindenstrauss theorem on projections of spaces, are related to Rough Sets
- But there are other relationships; it is natural to think about Rough Sets in terms of modal logic
- And, of course, there is a close relationship with Alexandrov topologies
- Close relationship with matroids, thus combinatorial optimization, was discovered

Envoi

- Certainly during this session there will be presentation stressing personal qualities of Professor Pawlak
- His multiple interests in science and elsewhere will be discussed
- Zdzisław was a true *Renaissance Man*. with contributions in many areas
- I was lucky (as were my contemporaries) to have such masters!