

2014

ATG Interviews: Raul Valdes-Perez

Katina Strauch

Against the Grain, kstrauch@comcast.net

Tom Gilson

Against the Grain, gilsont@cofc.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Strauch, Katina and Gilson, Tom (2014) "ATG Interviews: Raul Valdes-Perez," *Against the Grain*: Vol. 26: Iss. 4, Article 63.

DOI: <https://doi.org/10.7771/2380-176X.6910>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

CEO OnlyBoth and Founder of Vivisimo

by **Tom Gilson** (Associate Editor, *Against the Grain*) <gilson@cofc.edu>

and **Katina Strauch** (Editor, *Against the Grain*) <kstrauch@comcast.net>

ATG: Raul, you started your career on the Carnegie Mellon University computer science department faculty but from there went on to co-found Vivisimo, a company specializing in the development of computer search engines and now the new startup company OnlyBoth. Where did that entrepreneurial spark come from?

RVP: The sparks come from having an idea. It's no fun starting a company without a good idea. Luckily somebody invented computers and programming languages, so it's possible to achieve a broad impact by programming something novel and useful, without it being a massive undertaking. The ideas come from artificial intelligence, and my own focus is automated inference: figuring out what people do or could infer from data and background knowledge, and trying to automate the inference-making, or at least provide computational aids.

ATG: According to your press release, OnlyBoth has just launched "a novel technology that discovers new insights in data and writes them up in perfect English, all automated." How did you come up with such an innovative concept?

RVP: In late 1998, I was preparing for a family trip to Miami, Florida. As a research faculty member at Carnegie Mellon's computer science department, I thought to do something while there, so I visited the University of Miami's Website to see about giving a talk on our artificial-intelligence research. I had a revelation upon reading this sentence: "The University of Miami is the youngest of 23 private research universities in the country that operate both law and medical schools." Couldn't software be programmed to come up with such insights, which I began calling niches, and also write them up?

ATG: According to your Website, "this technology makes plain, in excellent English, what is hidden in masses of data." Is there a type data that is particularly amenable to this type analysis? What hidden relationships within the data is the technology most adept at discovering?

RVP: It works on structured data: entries (e.g., colleges, genes, members of congress, athletes, etc.) described by numeric, yes/no, symbolic, and even set-valued attributes. Part of the research challenge was identifying types of statements about individuals that people find interesting and understandable, and even partly inventing some new types, such as the rare "Only ... both" types of statements, which are quite rare in human practice, but are understandable, precise, and concise. Such statements gave rise to the company name.



ATG: We are obviously not computer experts, but as we understand it the structured data resides in fixed fields and is assigned a numeric, symbolic, or some other data attribute. Correct? In addition, there must be some technology that interprets the data and creates the readable English output, or are we missing something?

RVP: Yes, fixed fields, where the fields don't have free text, but a circumscribed range of values, e.g., income can have any number, birthplace can take on an uncounted number of names, etc. Of course, you have to tell the software what birthplace means, and a few other things about each attribute, but other than that, the software generates the insights and writes them up. Of course, humans programmed the software, just like teachers show students how to think and write, but are then not responsible for what the students do with this learning.

ATG: Structured data also includes data contained in relational databases and spreadsheets and depends on defining what fields of data will be stored and how that data will be stored. Can you recommend examples of data that you would like to work with?

RVP: The best applications for this technology are data sets where people care about individual entries (i.e., rows in the spreadsheet), and care about what makes an individual unique, surprising, or how an individual compares to its peers.

As a counter-example, suppose I have a dataset on daily physiological attributes of my own body over my lifetime. I don't care, nor does anyone else, what was special about me on May 7, 2007, unless aliens took over my body that day and some investigation is in order.

ATG: You've described OnlyBoth as a "reverse-Watson" — referring to the IBM supercomputer that gained fame on the quiz show Jeopardy. What did you mean by that?

RVP: My CMU doctoral classmate Oren Etzioni, now Director of the Allen Institute for AI in Seattle, was my house-guest and I was giving him a pre-launch demo of OnlyBoth Colleges. He remarked that it was like Watson in reverse, meaning that structured data comes in, and writings come out, unlike IBM's Watson and other ambitious text-mining projects, which take in writings and create structured facts and relations from there.

ATG: The inaugural OnlyBoth application analyzes data on 3,122 U.S. colleges and universities, but we understand that you will be expanding to create applications for other data sets. Can you elaborate?

RVP: Our immediate goal is to create awareness that this can be done, by launching applications that have wide appeal and which have available data, and to stimulate discussion about all the potential applications. Colleges are a natural first choice. We'll do sports next, and take it from there. You might call this a launch, lunch, and learn strategy.

ATG: We also wondered if you planned to add more datasets to the existing college application.

RVP: We certainly can if something interesting comes up. It takes minutes to add new data to the application. Then we push some buttons and wait a bit. My co-founder Andre Lessa, with whom I worked very closely on two technical projects at Vivisimo, has done a fabulous job of architecting the Web and mobile service and making it possible to run updates very easily.

ATG: Who is the intended market for the resulting applications? Is the focus on the individual subscriber or do institutions like libraries play a part in your marketing strategy?

RVP: Right now we're just launching free, impactful applications.

ATG: We can envision a situation where an institution like a library might want to create a database of easily readable output from data stored in their own repository. Is that a service that OnlyBoth would provide to interested libraries? If so, what is the process for making that happen?

RVP: We will listen to anyone who brings us interesting application ideas. We might not choose to address the opportunity, but we'll listen.

ATG: This is still a fairly new project. Have you developed a pricing model yet?

RVP: Right now we are just launching free, open applications.

ATG: It appears that OnlyBoth is very much in the formative stage. What are the

continued on page 40

next steps? How do you envision the future of the company?

RVP: We will launch free applications, grow awareness of this technology, learn from these launches and from discussions like this one, identify the best commercial applications, and go for them.

ATG: Given your work with *Vivisimo* and your experience in Web search and discovery systems, we would be remiss if we didn't ask you how you see search and discovery evolving. What can we expect in the near term? How about in the long term?

RVP: I'll let companies in the space talk about what's next. I can talk about what's needed. Users approach information retrieval systems with a variety of questions they want answers to. They then need to fit their need to the narrow channel of expressing a search query and getting back a list of results, with some navigation extras, perhaps. To address their need, users then need to do lots of extra manual and intellectual work.

Here's an example: Users often want to know what are the emerging topics in an area, i.e., what are emerging (or submerging) trends. You can do a query, but then you have to figure out yourself what the implicit trends are. **Andre Lessa** and I worked on that at *Vivisimo*, supported by a **National Science Foundation** grant, "Embedding Trend Discovery Within Search Engines," and got very promising results. As a user, I'd like search engines to offer to solve a variety of common information retrieval and analysis needs, not just search.

against the grain people profile

Chief Executive Officer, OnlyBoth LLC
P.O. Box 8120, Pittsburgh, PA 15217
OnlyBoth.com

Raul Valdes-Perez

BORN AND LIVED: Born in Havana, Cuba. Grew up in Chicago. Completed Ph.D at Carnegie Mellon and stayed in Pittsburgh.

EARLY LIFE: Attended elementary and high school in the Chicago area, then did a gap year in Madrid before college.

PROFESSIONAL CAREER AND ACTIVITIES: Got B.S.E. and M.S. in computer engineering at University of Illinois at Chicago. Then went abroad for two years seeking adventure. Returning to the U.S., worked at the MIT AI lab for two years in **Randy Davis's** group. Went to **Carnegie Mellon** to do a Ph.D in computer science, then stayed as a research faculty member. Finally, we started *Vivisimo*, which I served as CEO for nine years and chairman until **IBM** acquired us, 12 years after founding.

IN MY SPARE TIME: Family, friends, travel.

FAVORITE BOOKS: I often cite *Cyrano de Bergerac* as a romantic tale of heroism and nobleness, and *Brave New World* as a dystopia of a diminished mankind.

PHILOSOPHY: "Advice is for Winners" which is the title of my book, subtitled "How to Get Advice for Better Decisions in Life and Work."

MOST MEMORABLE CAREER ACHIEVEMENT: 15 years of monthly meetings with my Ph.D advisor and **Nobel Laureate Herbert A. Simon**, first as a student and later as a collaborator and colleague. One could ask him something about anything and walk away smarter.

GOAL I HOPE TO ACHIEVE FIVE YEARS FROM NOW: We want *OnlyBoth* to liberate the tremendous insights cloaked in structured data everywhere, yearning to breathe free and become known. 🍷

ATG: That sounds like a goal that we in libraries are trying to address. Instead of providing search results only, how do we provide possible meaningful solutions/pathways?

RVP Identify the common and valuable questions that users want answers to, and design algorithms and heuristics that will do it; it's a research endeavor. A further challenge is that you don't want yet another entry page that is separate from the search page. For our trend discovery work at *Vivisimo*, we added a separate Emerge button placed next to the Search button. A user typed a query and then pressed either Search or Emerge. Of course, you can't have too many such buttons, but you can have a few more. **IBM** now owns the prototype we developed. We'd love for them to bring it to market.

ATG: Everyone needs a little down time. What leisure activities do you enjoy? If we were to take a quick peek, what books would we see downloaded on to your tablet or smartphone?

RVP: I don't have a tablet or a smartphone, because I don't spend insufficient time in front of computer screens. For leisure, I prefer people contact involving friends, family, and travel, or books that help me see the world in ways that I'm not seeing now.

ATG: Raul, Thank you so much for a thoughtful interview! We really appreciate it.

RVP: You're very welcome. 🍷

against the grain publisher profile



OnlyBoth LLC

OnlyBoth LLC
P.O. Box 8120
Pittsburgh, PA 15217
Website: <http://onlyboth.com>

OFFICERS: Raul Valdes-Perez, Chief Executive Officer; Andre Lessa, Chief Technology Officer.

HISTORY AND BRIEF DESCRIPTION OF YOUR COMPANY/PUBLISHING PROGRAM: OnlyBoth discovers new insights in structured data and writes them up in perfect English. We say that only OnlyBoth does both. We also say that "A sentence is worth 1,000 data (TM)."

The idea for the technology came in 1998 and was supported by an NSF grant to Valdes-Perez while at Carnegie Mellon. Upon co-founding *Vivisimo* with a different technology, the idea sat on a shelf for 12 years. After **IBM** acquired *Vivisimo*, OnlyBoth co-founders Andre Lessa and Valdes-Perez saw that the idea remained novel and was even more relevant today, and so started OnlyBoth in March, 2014.

We seek to liberate the many insights cloaked in data by launching various applications for personal and business insights as well as fun. 🍷