Hilbert®

White Paper

A New Beginning

Version 1.0

Contents

INTRODUCTION 3

BREAKOUT OF "SAME KIND OF THINKING" 4

THE NEW PARADIGM 5

HILBERT ENGINE IMPLEMENTATIONS - PROVEN SUCCESSES 7

CUSTOMER EXAMPLE LEGAL SERVICES – MATCH IDENTITY 7

CUSTOMER EXAMPLE GOVERNMENT AGENCY – DOCUMENT EVALUATION 8

CUSTOMER EXAMPLE GOVERNMENT AGENCY – CASE AND RELATED DOCUMENTS 8

CUSTOMER EXAMPLE PHARMACEUTICAL – CLINICAL TRIALS 9

CUSTOMER EXAMPLE COMMUNICATIONS – MARKETING OPTIMIZATION 9

CUSTOMER EXAMPLE INDUSTRIAL SERVICES – DATA MINING MAINTENANCE DATA 10

CONCLUSION 10

ABOUT HILBERT 11

Introduction

The information age has evolved trying to analyze more and more data, yet we seem to understand less and less. Data mining for example requires dozens of people to interpret large amounts of data for many months without much of a result in terms of understanding the larger picture.

To analyze means to pull things apart and to scrutinize the individual pieces trying to determine the nature of the whole. The reality is that this can be done only if the system in question is simple enough to be fully comprehended by one or two minds. The number of parts in a larger system increases the complexity to such a degree, that it becomes difficult and eventually impossible to comprehend the whole.

Additionally, pulling things apart removes the many emergent properties that the aggregate dynamic system exhibits. For example, pulling a clock apart into its components and spreading the parts over a couple of shelves in a bookcase, will effectively remove the one emergent property that justifies the very existence of a clock, namely time. 'Time' cannot be found among the parts on the shelves. The same is true for the two gases Hydrogen and Oxygen, a most dangerous explosive mixture. However, when joined together they exhibit totally different properties. We discover a liquid called water, and it does not even burn. In fact, an opposite property emerges: it puts out fires.

Where do these emergent properties come from, and why can't we anticipate them from the individual components that make up the whole?

The answer to this question lies in the very nature of every system and the representation of its information: the whole is greater than the sum of its parts. The difference between the 'whole' and 'the sum of the parts' is called information and manifests itself as emergent properties. To understand such a system and to represent its information, it has to be conceptually simple. But for a system to be conceptually simple, it inevitably will be abstract. Because the more details we describe explicitly, the more complicated it becomes. On the other hand, the more details we formulate implicitly, the more abstract it becomes. Therefore, if we seek conceptual simplicity, we cannot avoid abstraction.

> ***"Every important idea is simple."***
> Leo Tolstoy in War and Peace

The explicit description of data is most often done by means of a natural language, such as English. On the other hand, implicit formulation of data can only be done by means of formal logic such as mathematics. If in this case one were to attempt the use of natural language, it would quickly lead to a nonsensical jumble of words.

Breakout of "Same Kind of Thinking"

It was in the early part of the last century that abstract tools were developed in the fields of Physics and Engineering that solved the most difficult problems in the understanding of natural systems that have ever been faced in the history of mankind.

Results from the field of Solid State Physics gave us the transistor, integrated circuits and computer chips, which produced an apparently endless stream of applications touching every facet of our lives. Results from Quantum and Relativistic Physics allowed man to split the atom and gave us atomic energy. Endeavors in Electrical, Chemical and Mechanical Engineering gave us rockets that took us to the moon and spaceships that traveled beyond.

> ***"We can't solve problems by using the same kind of thinking we used when we created them."***
>
> Albert Einstein

The magnitude of these problems and their solutions dwarfs any IT chore ever encountered. It stands to reason that with these tools the much simpler IT problems ought to be readily solved. However, one major obstacle stood in the way to allow this to happen. Virtually every IT undertaking has evolved in the last fifty years toward explicit descriptions of data and processes in natural language, mostly English text. In fact, databases and their myriad of tools together with their underlying languages have become gigantic word processors, and as such are fundamentally unsuitable for the abstract solution methodologies that allowed the enormous scientific and technological breakthroughs of the past.

The universe is not semantic, expressed in terms of morphemes, as our anthropocentric nature tries to make us believe. On the contrary, it is quantitative, expressed in numbers. Text does not lend itself to be operated on mathematically in the way numbers do. Therefore, it was first when a unique method of quantizing data was invented, a phenomenon never applied in the IT world before, that the door was opened to the full exploitation of these aforementioned tools and the success of the solutions to previously unsolvable problems became instantly apparent.

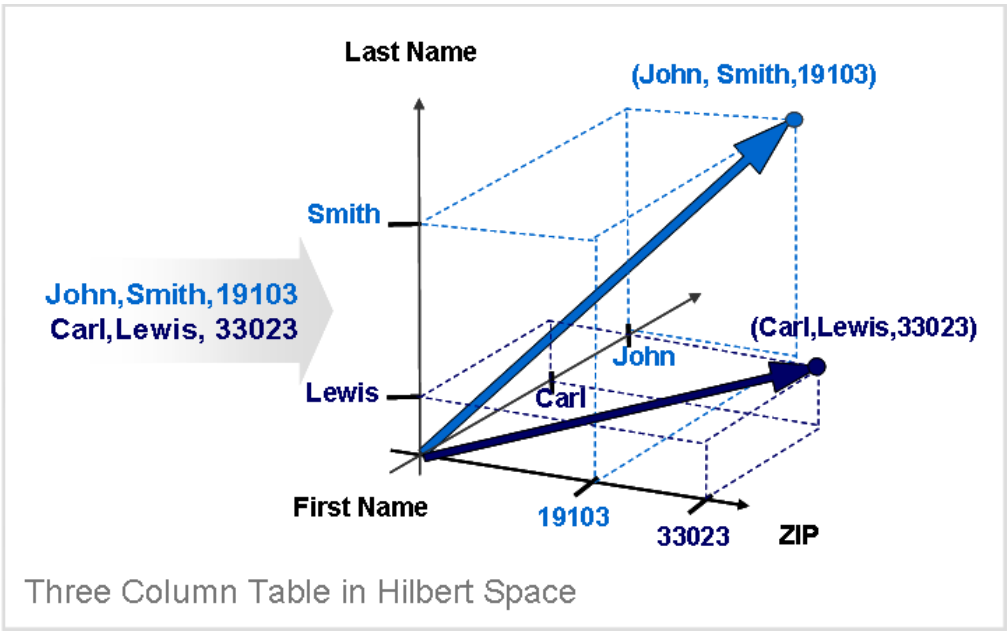
The New Paradigm

Once the gap to mathematical reasoning was bridged by quantizing the data, a whole new world of gaining information from large amounts of data at super speeds came into existence. Hilbert's software is based on its Hilbert® Engine, named after the famous German mathematician David Hilbert. All data are expressed as quantities via a unique process, protected by fundamental patents that have nothing to do with software. It allows related information from the fields of database records to be expressed as combinations of numeric entities, called mathematical vectors, while maintaining all semantic information of the data.

For instance, the simple combination of first name, last name and zip code would lead to a three-dimensional vector. Vectors have properties such as spatial direction and magnitude. Because most database tables contain considerably more than just three fields, the vectors will therefore also span much more than just three dimensions.

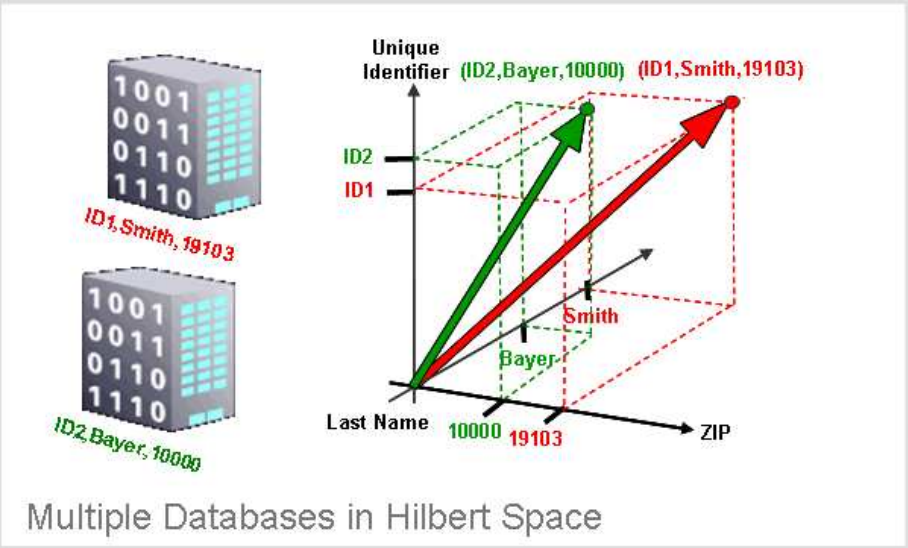
Each record, including all its relational fields, represents one vector and hence one point in a multidimensional space. Millions of records represent millions of points in that space – The Hilbert® Space as part of the Hilbert® Engine. After connecting the dots, a multidimensional body appears. This body of points is an interactive living representation of the actual system. There is no need to interpret on its behalf. It reacts to ad hoc queries and it speaks for itself. Simply put, one can pick and poke at it, and it reacts just as any system would react according to its own internal logic.

Large dynamic systems can now be represented mathematically rather than semantically. By combining and organizing the vectors, the Hilbert Engine can quickly find the most complex relationships. To be able to obtain, virtually instantaneously, complex relationships in context from huge amounts of data is far more efficient and goal oriented, than the endless and often mindless comparisons of individual database fields.



Three Column Table in Hilbert Space

To compare two or more disparate databases, the traditional wisdom prescribes the data content to be first normalized, a process in which large amounts of information will be lost. The Hilbert Engine however, provides space for each database. The resulting extended space allows the disparate vectors to operate in the same way all vectors operate. After all, an extended space is no more than a dimensional increase of an existing space. Therefore, all information remains intact and ready to be revealed in context and in real time.



Multiple Databases in Hilbert Space

Stated in simple terms, one can only seek and find knowledge in a holistic system, but one cannot find knowledge by looking at individual entities out of context, because one cannot see the forest for the trees. An automobile, totally disassembled with all its parts lying in different bins on the shelves in a warehouse, will be very difficult to climb into and then drive away. To turn the steering wheel found in one bin while looking at the wheels found in another reveals that the wheels don't move no matter how much one turns the steering wheel. The only relationship between the steering wheel and the wheels in the bins is the human interpreter, who usually is subjective, out of context and often just plain wrong, but most importantly, who can't make the wheels turn. On the other hand, when the parts are fully assembled into an automobile, nobody would argue about stepping into it and driving away. A functional automobile is a dynamic three-dimensional system. A corporation, an economic system, or a whole country, are no different. They are also dynamic systems albeit far more complex with many more parts and dimensions than an automobile. One can interact with them and observe the results.

The Hilbert Engine is not only the real-time solution to problems that till now have failed, but even more so, it provides the way to solve problems never before attempted and thus opens up new opportunities, new products and new markets. With the Hilbert Engine, the era of endless meetings of large groups of people trying to solve impossible problems at great expense have come to an end. Instead, proof of concepts or pilots can quickly be implemented with only a few people, at substantially reduced cost, that will readily illustrate the reality and performance of any given solution. Decisions can be made based on facts received from the results of the pilot rather than the promises of an elaborate report based on the lowest common denominator of the hopes and dreams of a committee.

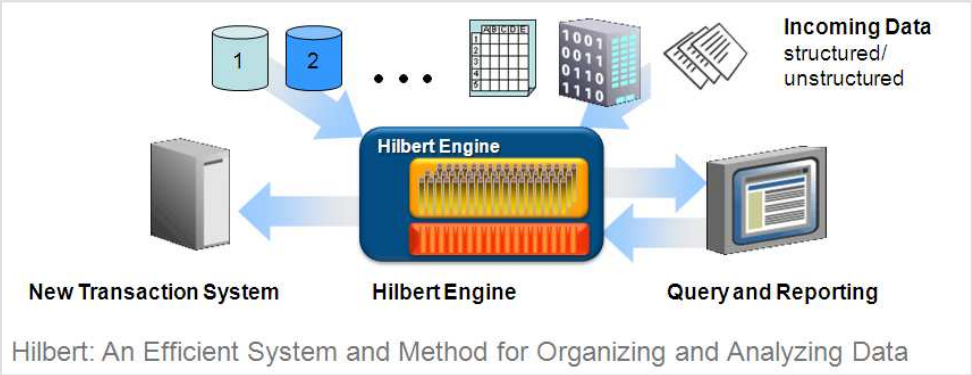
Hilbert Engine Implementations - Proven Successes

Here follow some practical examples of Hilbert Engine successes, where traditional methods failed.

Customer Example Legal Services – Match Identity

In one of the first applications of the Hilbert Engine, a legal services firm needed to match millions of financial records from more than fifty disparate data sources, each in its own peculiar format. The problem was that, besides format differences, much of the data content was incorrect because of data entry errors, manual preprocessing attempts, fraudulent alterations, etc.

Prior to Hilbert's involvement, one of the big five consultancy firms worked for several years on the problem to no avail. Subsequently, several others tried as well, but also failed.



Hilbert: An Efficient System and Method for Organizing and Analyzing Data

After having spent millions of dollars on these futile activities, as a last attempt, the new Hilbert technology was applied.

1. **Import of Database Data** into Hilbert Space
2. **Quantization of Data** to establish a multidimensional infrastructure in Hilbert Space
3. **Query of Data** to gather vectors of interest
4. **View-Vectors** were explored and compared
5. **Record Identification Numbers** were automatically generated in an iterative analysis process

The outcome was that the Hilbert Engine matched more than six million financial records in less than two minutes on a laptop. The process included over thirty million related records, with an average of three hundred attributes per record. The records were matched to an accuracy of one out of twenty billion. Critical financial reports that previously required more than twenty hours to produce are now generated in a few minutes.

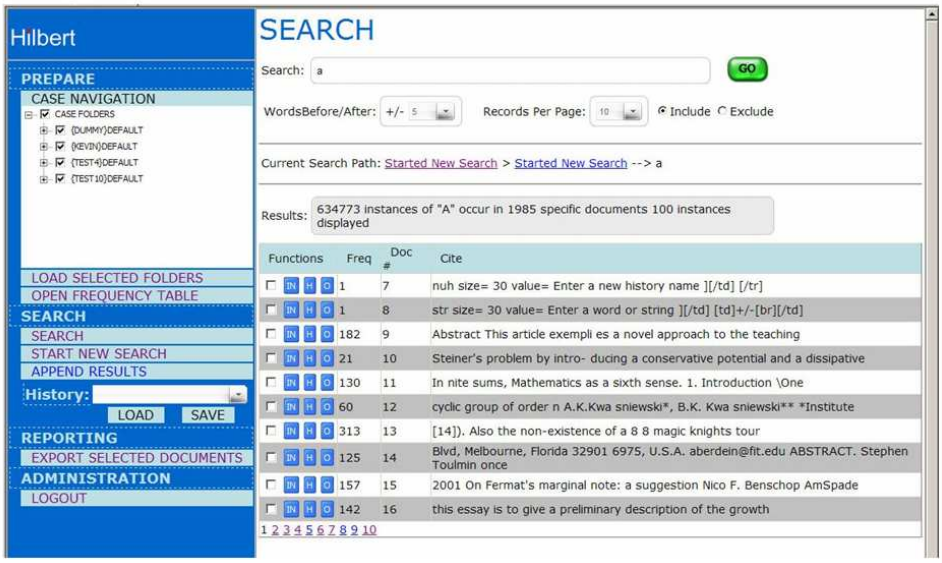
This is one example of how the Hilbert technology is all about fast, practical and meaningful solutions at reduced cost and not about long-term social gatherings resulting in huge reports that explain why things cannot be done.

Example Government Agency – Document Evaluation

To combat the threat of terrorism, our government agencies are searching for better tools to automatically evaluate large quantities of documents and messages in seconds. For a DARPA (Government Agency) project, Hilbert used the Hilbert Engine to search tens of thousands of unstructured documents with virtually instantaneous response to any query. This is an interactive investigative process that outperforms any commercially available search engine in finding “needles in haystacks” in any type of document, email message, voice transcription or other text-based content. The outcome is both speed and accuracy.

Example Government Agency – Case and Related Documents

The DOJ (Department of Justice) is currently evaluating the Hilbert Engine to increase their ability to investigate hundreds of millions of pages of unstructured content in support of a potential case. Additionally, the system also automatically correlates documents as to subject matter in just seconds. Again, the outcome is both speed and accuracy, and the DOJ lawyers will now be able to dramatically improve their productivity.

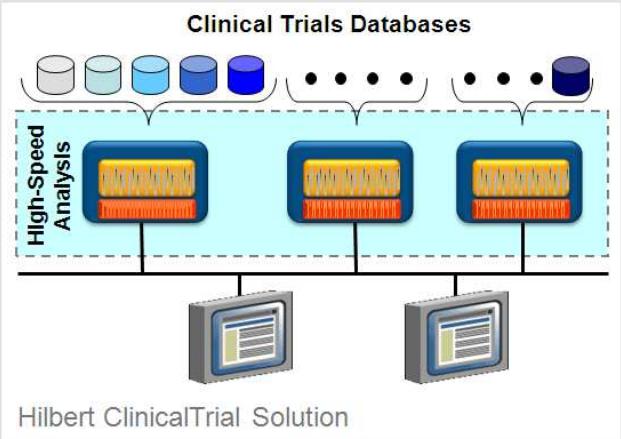


Hilbert Case&Documents Management

Example Pharmaceutical – Clinical Trials

In the pharmaceutical industry, understanding the information found in clinical trials data is critically important. Each drug requires many individual trials, which can take months, even years to complete. They can contain enormous amounts of data in various formats that include hard facts as well as the doctor's written observations. Due to the format differences and the data complexities, one trial is almost never compared to another, thereby loosing the ability to see the larger picture.

One major pharmaceutical company applied the Hilbert Engine to perform analysis across sixteen different and independently run clinical trials. By quantizing all the data from each clinical trial into individual subspaces, which together make up one multidimensional vector space, the data can now be analyzed universally. Much of the critical information resides in the unstructured observational notes of

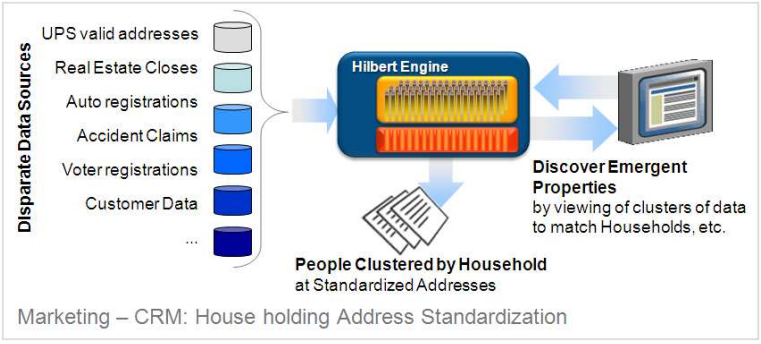


Hilbert ClinicalTrial Solution

the doctors running the trials. With the Hilbert Engine, this pharmaceutical company can now put both the structured as well as the unstructured data into context, and discover relationships never before revealed.

Example Communications – Marketing Optimization

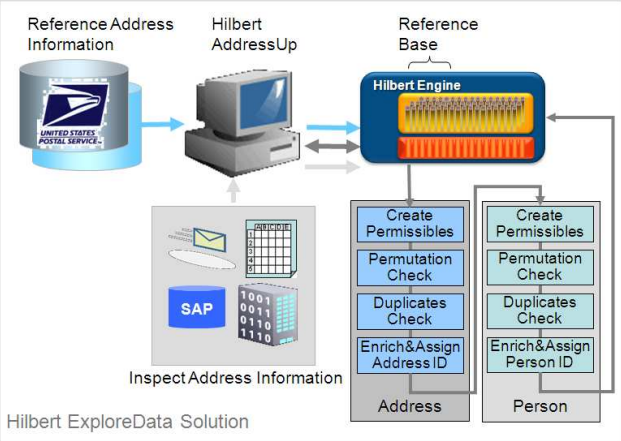
Large telecommunication companies have their own sets of data problems. These companies spend hundreds of millions of dollars each year in an effort to grab their competitors' customers, which require data from other sources. These sources contain many duplicates, but also repetitions of their own customers.



Marketing – CRM: House holding Address Standardization

Poor quality data costs these companies millions of dollars in duplicate mailings, missed prospect opportunities, upset customers, etc. Customer matching systems are deployed in order to identify unique addresses, eliminate duplicate records, and to classify the residents at each address into one or more households. These systems are built using conventional databases and typically use hundreds of thousands of semantic rules in an attempt to keep the data clean. They are at best sixty to seventy percent accurate.

The Hilbert Engine application that is not based on semantic rules but on mathematical logic. Instead of processing rules, the Hilbert Engine first quantizes each address in the customer and prospect databases, and compares them vectorially to a two billion record, quantized reference base of valid addresses from the US Postal Service. This corrects the address por-



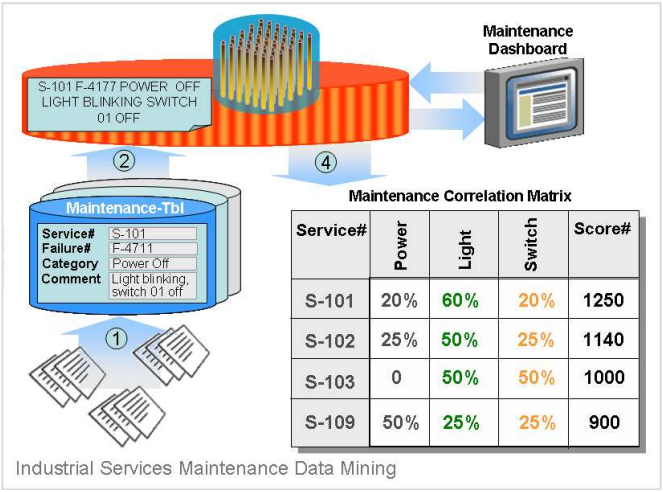
Hilbert ExploreData Solution

tion of the name/address vectors. Now the complete name/address vectors are compared with one another to produce households. Hilbert's AddressUp application achieves ninety five percent accuracy. It processes many more names and addresses per unit time, using a personal computer, than any other system can on mainframes.

Customer Example Industrial Services – Data Mining Maintenance Data

Aircraft system failures are a growing problem. Large teams of skilled engineers, seeking to predict failures, tried for many years to analyze maintenance data. The results were limited and certainly not real time. Additionally, these teams attempted to use numerous commercial data mining tools, which all failed. By leveraging its ability to solve problems involving both structured and unstructured data simultaneously, the Hilbert Engine demonstrated that it does the job in minutes.

The initial skeptics became Hilbert's biggest supporters.



Service#	Power	Light	Switch	Score#
S-101	20%	60%	20%	1250
S-102	25%	50%	25%	1140
S-103	0	50%	50%	1000
S-109	50%	25%	25%	900

Industrial Services Maintenance Data Mining

Conclusion

These examples, as well as others, prove that the Hilbert Engine represents a breakthrough in information science. It solves problems that conventional systems cannot. Information arises from data when it is assembled in context, and the Hilbert Engine represents the only technology to achieve that assembled state. The door has finally been opened to an exciting future of discovery.

About Hilbert

Pennsylvania based Hilbert Technology Inc. is an international provider of data management and analysis solutions for large and medium size enterprises worldwide. The offering is based on the revolutionary, patented Hilbert Engine technology for the ultra high-speed access, manipulation, storage and analysis of large volumes of structured and unstructured data. The Hilbert solutions are offered as industry or process specific solutions. Organizations in public services, law, government, finance, communications, whole- and retail sales, transportation & tourism and chemical & pharmaceutical can employ Hilbert solutions to gain unparalleled speed in access and analysis over large data volumes.