

Beyond Knowledge Discovery

Effective and timely analysis is critical to a company's overall success and competitive position. At some level, all business process owners utilize analytics to more effectively evaluate options, determine strategy and set goals. Business analytics provide the foundation for decisive decision-making in response to dynamic markets and business conditions. But more than enabling a smarter reaction to unwelcome surprises, PolyVista Enterprise Analytics offer proactive and early warning results that can put you on the offensive. By providing a seamless integration of analytical techniques and capabilities, PolyVista uniquely delivers both answers and insight to business users across the enterprise.

PolyVista analytical software was designed to support and enhance the valuable process of *Interactive Knowledge Discovery* (IDK). In stark contrast to canned reporting and managed query, IDK can surface new and valuable business opportunities or bring to light previously unknown operational anomalies. It is just the type of analysis required to fuel "Out of the Box" thinking and enable companies to redefine their markets and products. This interactive discovery process is characterized by very fast query response, intuitive visualizations and automated discovery techniques. The PolyVista analytical environment brings business users together with their data and aids them in discovering new and valuable "Business Insights" hidden deep within the data. While this pro-active style of Knowledge Discovery is still considered revolutionary in the context of today's business enablers, this paper will extend this important concept and describe the next level of empowering knowledge workers.

Robert Ford
PolyVista Inc
VP Technical Services

What do you Know?

For the most part, successful companies out-perform their competitors by being more efficient and innovative in executing their plans. In that most companies operate largely on human capital (knowledge), it only makes sense that leveraging this knowledge is critical to a company's overall success. Relevant experience and insights gained over time are extremely valuable and these knowledge workers should be an integral part of a company's decision making and planning process. The question becomes, are you leveraging the collective knowledge of these individuals and applying their expertise to innovate and improve your business processes?

One way to describe various "states of knowledge" in relation to analytics is represented in the following matrix;

State 1- "*Knowing what you Know*" is a situation generally supporting "business as usual". Here, the business fundamentals are well known, specific relationships are understood, operational metrics have been identified and are being measured, tracked and reported. Reporting & Publishing is a predominant activity. This state can induce a "false" sense of security.

	Know	Don't Know	
Know	State 1 Know what you Know Standard Reporting	State 2 Don't Know what you Know Knowledge Management	Know
Don't Know	State 3 Know what you Don't Know Interactive Query	State 4 Don't Know what you Don't Know Discovery (data mining)	Don't Know

State 2- "*Don't Know what you Know*" is the classic case of Knowledge Management where the right-hand doesn't know what the left-hand is doing. It describes a company that is not effectively communicating or making "known knowledge" available to those who need to know it. In the area of analytics the relevant question is, how do business analysts and business managers leverage each other's experiences, discoveries and techniques?

State 3- "*Knowing what you Don't Know*" describes a condition where it has been recognized that changes are required to sustain the business. The status quo is not acceptable. Problems and/or opportunities are known or suspected but a "lack of information" exists to solve them. Given that the problem area is defined, On-Line Analytical Processing (OLAP) is a very effective technique to understand and quantify the solutions.

State 4- "*Don't Know what you Don't Know*" is a very interesting yet precarious situation to be in. In business, what you don't know *can* hurt you! Being unaware of issues and opportunities will adversely affect your company's performance and worse, your ability to effectively compete.

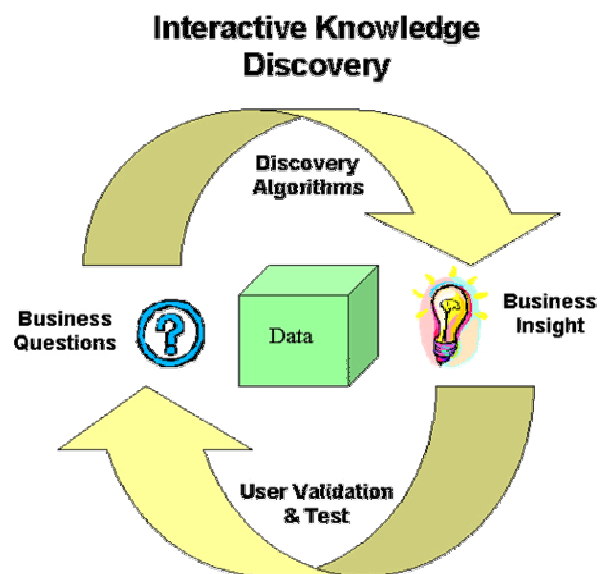
Only through Discovery techniques can potential new knowledge or "insight" be gleaned and then put to use. This new knowledge may represent a new marketing opportunity or

perhaps surface a previously unknown business process improvement. Without a means to Discover these insights, company's will not only miss valuable opportunities but also become vulnerable to more pro-active competitors.

PolyVista's unique and innovative Analytical Solutions deliver value in all quadrants of this knowledge matrix. Ad-hoc reporting and interactive analysis are delivered via PolyVista's "Best in Class" OLAP browser functionality. Collaboration and knowledge sharing are supported through the PolyVista Wisdom suite, and our Discovery algorithms address the answers to questions not yet asked.

PolyVista delivers Interactive Knowledge Discovery today

PolyVista empowers analysts and business professionals to leverage their knowledge and experiences in a guided search for new and valuable business insight or opportunities. By seamlessly integrating the rich functionality of On-Line Analytical Processing (OLAP) directly with interactive 3D graphics and the automation of Discovery Algorithms (data mining), PolyVista provides a "discovery focused" experience for the end user. Our Discovery algorithms (data mining) are designed for simple configuration and are tuned to leverage high performance OLAP databases. These multi-dimensional databases provide for very fast query response, ideally suited for users generating ad-hoc queries as well as for supercharging our Discovery Algorithms. These algorithms can surface Business Insights (ie *nuggets*) that represent discovered anomalies, relationships, or business rules hidden deep within the OLAP database. These insights may take the form of a better understanding of customers and their buying patterns; they could also represent new business opportunities or may reveal anomalies and inefficiencies in business processes. The discovered insights are then presented to the user for their business value assessment. Many times these insights represent important actionable business knowledge or they may form the basis for further analysis and verification, all seamlessly supported by our tools. We refer to this process of discovery and verification as "Interactive Knowledge Discovery". Each iterative cycle leverages the business user expertise and the computers processing speed, resulting in a very efficient search for actionable Business Insight.



Compare PolyVista's approach of automatically surfacing *nuggets* to other manually driven query tools and report writers on the market. These query tools all require the user to formulate various questions (or queries) to be answered from the database. If the user formulates an irrelevant question, then the result will also be irrelevant ("Garbage In / Garbage Out"). Also consider that a user after a few weeks of

struggling with a simple query tool, will no doubt exhaust their personal list of questions to answer. At best, they get answers to those questions they are able to correctly formulate, but at the end of the day little new knowledge or insight is uncovered. PolyVista provides the best of both worlds, a best in class slice & dice query tool but more importantly, the power of automating the search for new knowledge hidden away in the database. Thus, PolyVista proactively helps users ask better questions which lead to more insightful answers.

Empowering Knowledge Workers

Competitive companies cannot solely rely on a handful of backroom analysts to examine data for trends and anomalies or to identify new business opportunities. In fact, without a working knowledge of the business, many high-powered analysts while brilliant statisticians could easily overlook an important business insight or opportunity. Therefore these companies are equipping their knowledge workers at all levels with powerful yet easy to use tools that will leverage the users business know-how and experiences. Each user can easily configure PolyVista to search for interesting relationships or anomalies and then interactively review the discovered *nuggets* for relevance and value. While PolyVista is naturally powerful in the hands of trained business analysts or statisticians, the greatest value to a company is positioning PolyVista to enhance the collective effectiveness of its decision makers and planners. This larger group of knowledge workers has both the need for timely analysis as well as the business experience and subject matter knowledge to identify potentially valuable nuggets. It is this group that can quickly validate and “act on” important Business Insights or Opportunities that PolyVista brings to light. It is clear that PolyVista is well suited to enhance the capabilities and value of individual knowledge workers, but how does that scale to an Enterprise Solution?

Putting Enterprise Knowledge to Work

Building upon PolyVista's unique abilities to raise the collective value of a pool of individual knowledge workers, PolyVista-Wisdom (PV-Wisdom) defines the next generation of Enterprise Analytics. PolyVista-Wisdom is our scalable Enterprise Analytics solution designed to deliver qualified, business relevant insight to any user... anywhere.

PV-Wisdom's server based architecture extends PolyVista's Interactive Discovery model to include an even broader base of a company's knowledge workers. In this model, users can be completely oblivious to the intricacies of data mining and are simply offered the fruits (results) of the data mining labor. In addition, our architecture supports the capability for specific results to be tagged (scored) as particularly interesting or valuable. These "pre-qualified" results can then be searched for by end-users or even delivered to them for assessment or action. The PolyVista Discovery Server can be configured to run continuous and/or batch jobs that keep the Knowledge Repository populated with the most current results. Thus, even users who have no interest or desire to configure or directly apply data mining algorithms can still reap their benefits.

While PV-Wisdom can be thought of as a Knowledge (*nugget*) repository, it has been designed and architected to manage the critical processes involved in finding and delivering valuable and relevant business insight to decision makers across an organization. These generic processes include;

- Discovery – finding nuggets using multiple analytical algorithms and techniques
- Result Validation & Context – tagging or qualifying nuggets for further investigation, adding context/comments to nuggets sometimes involving multiple contributors, detailed analysis of value
- Intelligent Search & Retrieval – The PV-Wisdom repository is accessed using a search engine metaphor and provides for dynamic user profiling that leads to more relevant search results

Discovery

The server Discovery process is powered by PolyVista's server based, high performance analytical engine. This server is configured to run specialized Discovery Algorithms (data mining) that direct their results to the Wisdom Repository or Knowledge Base. Using this centralized approach, a small number of trained analysts can support the mining needs of the larger population of knowledge workers who have no desire to configure or tune Discovery Algorithms. For those analysts who are inclined to configure algorithms for specific objectives, the Wisdom repository can also accept mined results (nuggets) directly from these power-users. These analysts can publish selected results to the nugget repository that they have found particularly interesting and potentially valuable to other business analysts.

Our server based architecture supports intelligent scheduling, load balancing and multiple client interface options. Numerous algorithms are available that perform generic statistically based searches for trends and anomalies, or others targeted for specific pattern classification or prediction. Supported algorithms include;

- Deviation
- Summary

- Market Basket Rules
- Association Rules
- Decision Rules
- Probability Rules
- Clustering

Validating Results and adding Context

The Validation process highlights specific results (stored in the Wisdom Repository) as candidates for further investigation and value assessment. This Qualification Role could involve technical users who are very familiar with specific business areas and the data that support those processes. The role includes reviewing the knowledge base for new and potentially interesting results and validating those prospective nuggets to ensure that they are relevant and accurate. In that many of the nuggets will have been identified through server based mining, these individuals need only have strong business knowledge and not be data mining experts. The Wisdom repository has been designed to allow each nugget to be assigned numerous attributes. These attributes not only include the basic result facts such as relevant dimensions, measures and algorithmic scores, but more importantly it allows users to add textual commentary and assign an overall rating of value or significance as well. These rich nugget attributes enable a robust search method where users can easily constrain their search of the knowledge base to relevant nuggets of interest.

Assessment & Collaboration

Once nuggets of a general interest have been identified, a more formalized assessment process may be initiated to transform a *nugget* from an “interesting anomaly or relationship” into one or more actionable tasks that could be tracked to completion. The nugget is reviewed to such a level that a business case can be created to justify the proposed issue, solution, costs and benefits. Many times collaboration is very helpful in fully exploring the issues and benefits related to a certain result or someone’s interpretation of that result. In this way the knowledge repository serves as a share-point enabling users to collaborate on related nugget issues such as result validity, relative value or “appropriate next steps”. Comments and discussions relevant to the analysis and subsequent course of action can also be captured. To close the knowledge loop, learning’s from discovered nuggets and related collaboration is then used to fine tune or even redirect the server-based algorithms. While apparently describing a cyclical process, it is more like a spiral where each full iteration not only completes the cycle but also advances to a higher level of business knowledge and awareness.



Summary

PolyVista Wisdom represents a unique, innovative and scalable Enterprise decision-support solution. As an integral part of a company's decision-making processes, PolyVista enables business users to respond to issues and opportunities both seen and unseen, known and unknown. The PolyVista Wisdom architecture extends the reach of data mining benefits to the non-analyst. Centrally managed server based algorithms direct their results (nuggets) to the Wisdom database allowing users to search/subscribe-to the repository for relevant results. Direct support for capturing commentary and ratings enable a collaborative environment to provide important context to otherwise basic facts. Thus we are able to effectively deliver new "Business Insight" not only to business analysts and researchers, but also to managers and decision makers at all levels. Actionable results can then be "transferred" from the Wisdom environment to perhaps a Knowledge Management or Task Management system for follow-through.

Through PolyVista's proactive forward-looking analytics, businesses will not only respond decisively to a competitor's challenge, but more importantly position itself to "set the bar" for its competitors.

Appendix to “Beyond Knowledge Discovery”

Comparing OLAP, Discovery and Wisdom (an illustrative example)

To help illustrate these concepts, let’s walk through a couple of analytical scenarios to compare the various analytical techniques. For these cases, let’s assume we are in the grocery store business with numerous store locations and store types. Also assume that we have created a multi-dimensional Sales database (cube) from data captured in our Point of Sale systems and related Customer databases.

Slice & Dice

Using this Sales Cube and the PolyVista OLAP browser, one can easily do some ad-hoc reporting and slice & dice analysis. Common queries may be to show Profits across all Food Categories by Quarter, sorted by Yearly Totals.

		1997	1997			
			Q1	Q2	Q3	Q4
Food	Produce	49,417.09	12,339.83	11,590.31	12,374.95	13,112.00
	Snack Foods	40,646.48	10,274.58	9,434.91	10,408.67	10,528.31
	Frozen Foods	33,176.84	8,171.63	7,798.58	8,273.95	8,932.68
	Canned Foods	23,879.81	5,912.34	5,592.34	5,870.63	6,504.49
	Baking Goods	23,299.80	5,773.45	5,469.27	5,743.82	6,313.26
	Dairy	18,280.00	4,612.52	4,338.83	4,433.32	4,895.32
	Deli	15,210.06	3,747.85	3,503.46	3,635.26	4,323.49
	Baked Goods	9,891.34	2,414.97	2,189.63	2,493.96	2,792.78
	Snacks	8,722.47	2,137.96	1,884.16	2,279.58	2,420.78
	Starchy Food	7,050.16	1,790.50	1,692.31	1,682.05	1,885.30
	Eggs	5,515.86	1,168.56	1,348.14	1,415.35	1,583.81
	Breakfast Foo	4,184.66	1,030.98	976.42	1,029.26	1,148.00
	Seafood	2,288.44	510.71	517.13	548.20	712.40
	Meat	2,204.47	495.89	509.25	586.94	612.39
	Canned Produ	1,997.39	432.69	478.62	486.60	599.47

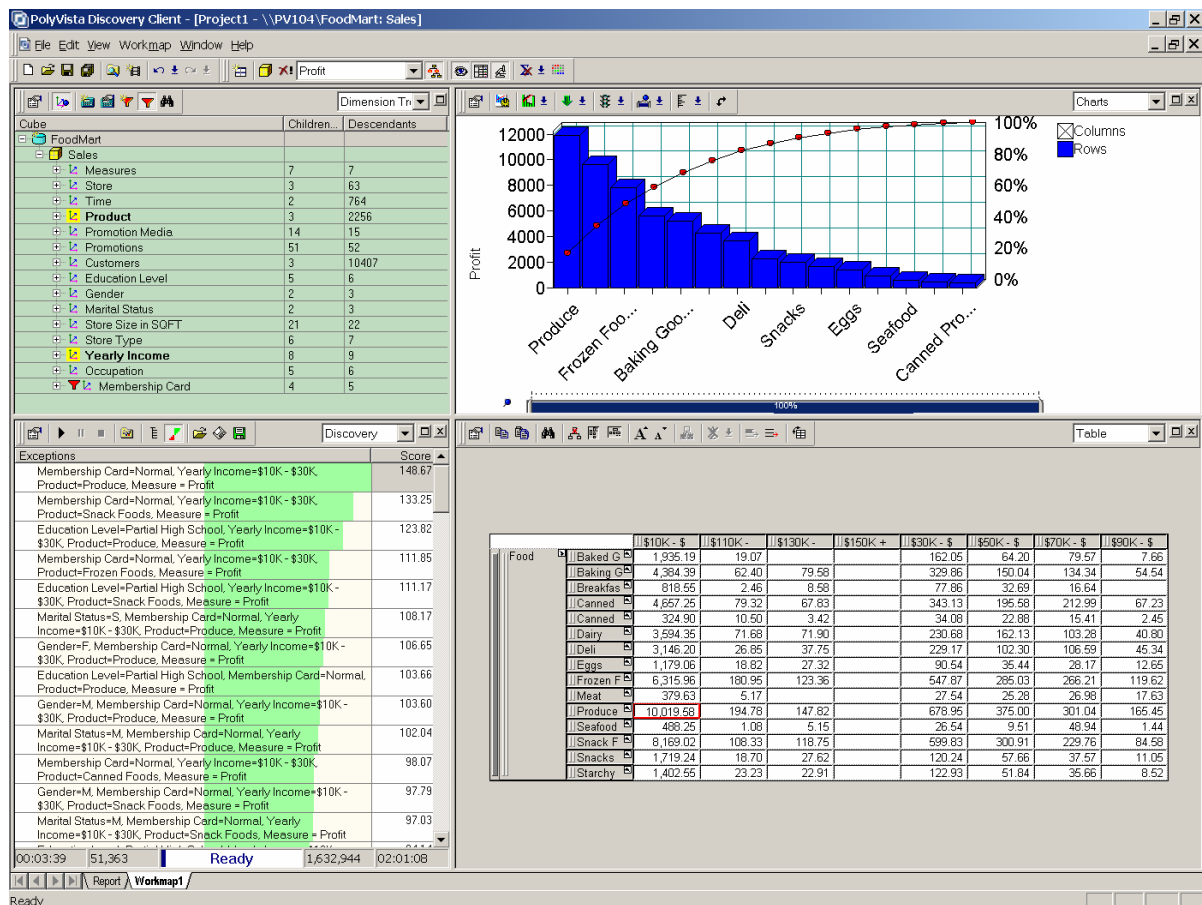
Other dimensions we could use to manually slice various sales related measures are Store level geography, Promotions, Customer geography and Customer demographics like gender, salary range, education level, occupation, etc. In addition to the basic OLAP functionality of slice & dice, we can also “drill down” to various levels of detail in the data, display data using charts and graphs, sort rows or columns, display totals, etc. This type OLAP functionality is most useful when you know or have a good idea of what question you are asking of the data. For example, in the query we defined previously, we knew explicitly what data elements we wanted to include in our table or report. In other cases, the task is far more complex. What if the task was to find or discover a new trend, pattern, or anomaly relating Profit to any or all dimensions in the cube. Or how would one find a profile that identifies a highly profitable or highly un-profitable customer? Given the speed at which OLAP delivers results to my manual queries we are tempted to claim that we can just slice & dice our

way to the answers. In this case an OLAP strength becomes its weakness. In other words, the very benefit that OLAP delivers in enabling us to view the Profit measure across any Dimensional combination and hierarchy, becomes a daunting if not impossible task when you consider all the possible combinations that would have to be explored.

Knowledge Discovery with OLAP & Data Mining

Consider now the huge advantage in imbedding Discovery techniques (data mining) with OLAP. In this model, we are able to configure a Discovery Algorithm to assist us in “exploring” a cube for interesting relationships or anomalies.

The following image depicts the results of the PolyVista Discovery algorithm where the search is for Profit anomalies (Profit cell values that are significantly greater than or less than an expected Profit value for that cell) across all combinations of 8 different dimensions and/or dimensional hierarchies.



As the algorithm runs, it will begin to populate the Analysis Result Pane located in the bottom left-hand portion of the PolyVista application. This list represents the top fifty (positive scores) and bottom fifty (negative scores) most statistically interesting results. The list is sorted such that the greatest positive scores are at the top and the most negative scores are sorted to the bottom of the list. In this example, the most positively interesting result (a profit value significantly greater than expected) is

identified as customers who use the Normal membership card, make between \$10k - \$30k salary and purchase Produce.

Given this sorted list of anomalies, we can review the results looking for one that catches our attention as being either curiously interesting or perhaps completely un-intuitive. In any case, we select the result to explore further, click on it and drag it to the table pane where the query is immediately executed. Now with all the OLAP functionality at our disposal, we can begin to answer the many follow-up questions such as what products are these customers buying, where do they live, what promotions do they respond to, etc. Basically we have taken a *nugget* that the PolyVista Discovery Algorithm had surfaced, and embellished that result with enough follow-up analysis to perhaps develop a new target marketing campaign. In other cases, the Discovery algorithms may have surfaced an interesting relationship and my validation process would provide enough new information to fine-tune the algorithm for another iteration of exploration/validation.

This example illustrates the power of seamlessly combining OLAP with automated mining techniques and putting this capability in the hands of your knowledge workers.

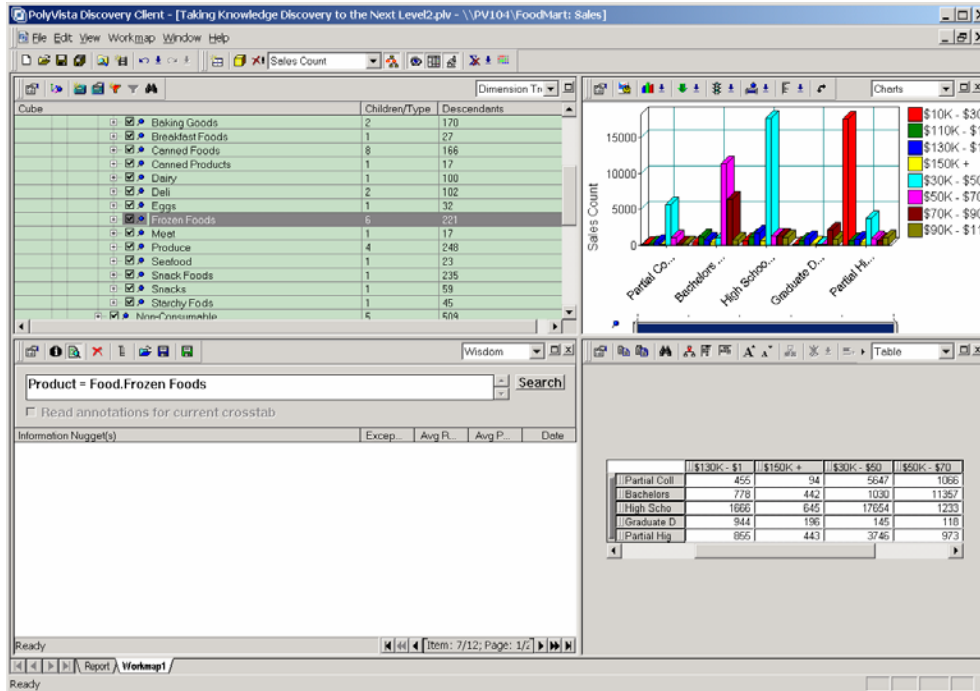
Knowledge Discovery - Wisdom

In the previous example, we saw how a business analyst using PolyVista Analytical Client might engage in a Knowledge Discovery process. This involved configuring a Knowledge Discovery algorithm, reviewing and validating results, and perhaps re-tuning the algorithm for a more specific search and repeating the process. As we consider scaling the use of PolyVista technology to an Enterprise level, we acknowledge that not everyone will be interested or capable of configuring or fine-tuning algorithms and validating their results. In fact the majority of potential beneficiaries PolyVista technology will simply want qualified *nuggets* or the *fruits* of the data mining labor.

The PolyVista Wisdom architecture has been designed to support and enhance this nugget focused analysis process.

In keeping with the previous example, the Wisdom architecture would include a process where a mining administrator has configured a number of server-based algorithms to run against the Foodmart Sales database. The results of the analytics (*nuggets*) are continuously directed to our Wisdom SQL database. The Wisdom DB acts as a *nugget* repository and not only manages the results from multiple algorithms but also tracks important attributes associated with each result. Thus this architecture supports a wider class of business users who simply “search” the Wisdom DB for relevant nuggets (business insights) rather than have to deal with any data mining algorithms. This “search” process would be modeled after internet style search engines where keywords and other user defined criteria limit the set of *nuggets* returned to the user for their inspection.

The bottom left panel of the following screen-shot depicts a prototype of the search screen interface for Wisdom..




In the previous screen image notice that the search text box has been populated with the term *Product = Food.Frozen Foods*. This key word string was the result of dragging and dropping that same named member from the highlighted Products dimension tree (above it). Pressing the Search button runs a query against the Wisdom Repository (nugget database) returning mining results that include the listed keywords and satisfying parameters defined in the users search profile. The next image shows the result of the search action.

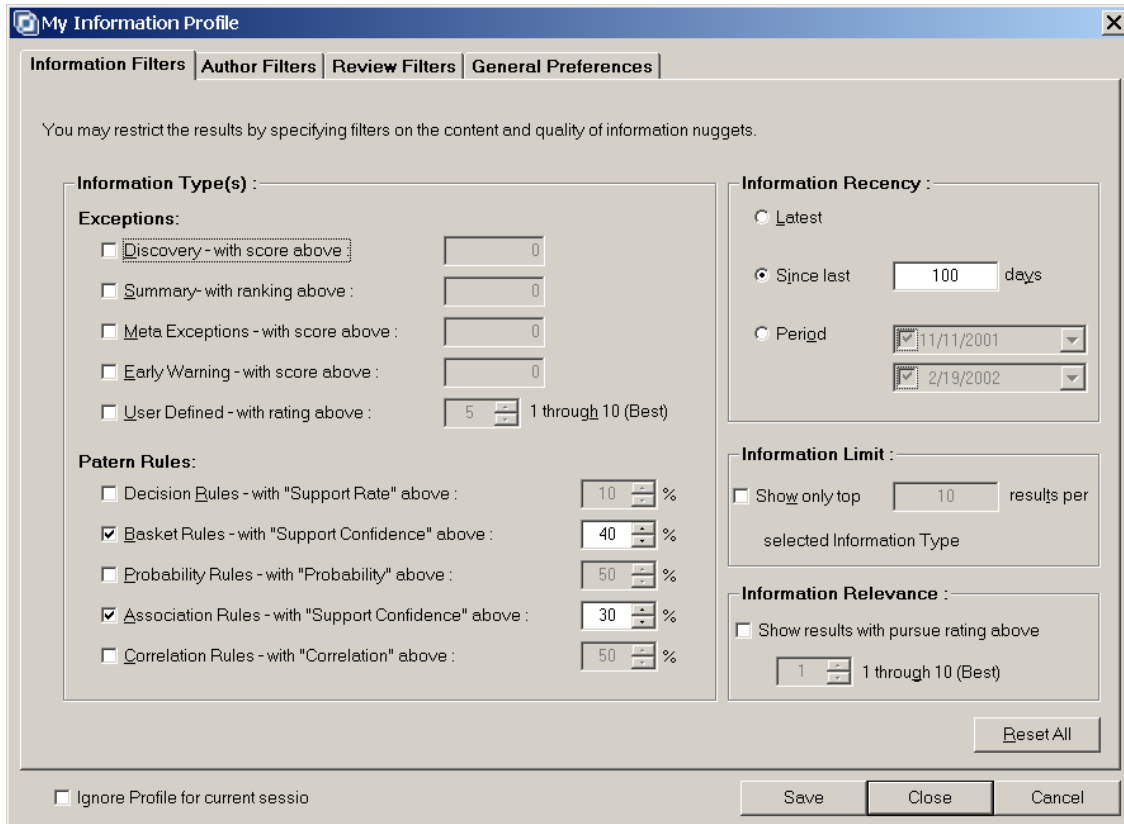
The screenshot shows the PolyVista Discovery Client interface with the search results for 'Product = Food.Frozen Foods'. The search bar contains the text 'Product = Food.Frozen Foods' and a 'Search' button. Below the search bar, there is a section for 'Information Nugget(s)' with columns for 'Excep...', 'Avg R.', 'Avg P.', and 'Date'. The results are displayed in a table with the following structure:

Information Nugget(s)	Excep...	Avg R.	Avg P.	Date
Description & Results				
Description				
Generated Tuples				
Basket Rules				
Product = Food.Frozen Foods, Product = Food Produce, Product = Food Snack Foods	45.16	0	0	2/20/2002 6:44:58 AM
Product = Food.Frozen Foods, Product = Food Produce, Product = Non-Consumable Household	43.2	0	0	2/20/2002 6:44:58 AM
Product = Food.Frozen Foods, Product = Non-Consumable Household, Product = Food Snack Foods	45.44	0	0	2/20/2002 6:44:58 AM
Product = Food.Frozen Foods, Product = Food Snack Foods, Product = Food Produce	47.97	0	0	2/20/2002 6:44:57 AM
Product = Food.Frozen Foods, Product = Food Snack Foods, Product = Non-Consumable Household	43.95	0	0	2/20/2002 6:44:58 AM
Product = Food.Frozen Foods, Product = Food Produce	47.19	0	0	2/20/2002 6:44:56 AM
Association Rules				
Product = Food.Frozen Foods, Membership Card = Bronze	30.65	0	0	2/19/2002 7:28:08 AM
Product = Food.Frozen Foods, Gender = M	28.08	0	0	2/19/2002 7:28:06 AM
Product = Food.Frozen Foods, Yearly Income = \$30K - \$50K, Education Level = High School Degree	32.61	0	0	2/19/2002 8:46:29 AM
Product = Food.Frozen Foods, Marital Status = S	28.08	0	0	2/19/2002 7:28:07 AM

At the bottom of the window, there is a status bar showing 'Connected.' and navigation controls.

Nugget results can now be reviewed based on the rule's member list and rating metrics. Ratings include both a composite score for rule significance and accuracy, as well as ratings that reflect "Reviewers" personal assessments too. As with all Mining results, interesting nuggets can be dragged over to the PolyVista's table pane to review and validate.

The user's "Search Profile" is accessed via the  button on the Wisdom tool bar. The next image depicts the Profile dialog panel.



My Information Profile

Information Filters | Author Filters | Review Filters | General Preferences

You may restrict the results by specifying filters on the content and quality of information nuggets.

Information Type(s) :

Exceptions:

- ☐ Discovery - with score above : 0
- ☐ Summary - with ranking above : 0
- ☐ Meta Exceptions - with score above : 0
- ☐ Early Warning - with score above : 0
- ☐ User Defined - with rating above : 5 1 through 10 (Best)

Pattern Rules:

- ☐ Decision Rules - with "Support Rate" above : 10 %
- ☒ Basket Rules - with "Support Confidence" above : 40 %
- ☐ Probability Rules - with "Probability" above : 50 %
- ☒ Association Rules - with "Support Confidence" above : 30 %
- ☐ Correlation Rules - with "Correlation" above : 50 %

Information Recency :

- ☐ Latest
- ☒ Since last 100 days
- ☐ Period: 11/11/2001 to 2/19/2002

Information Limit :

☐ Show only top 10 results per selected Information Type

Information Relevance :

☐ Show results with pursue rating above 1 1 through 10 (Best)

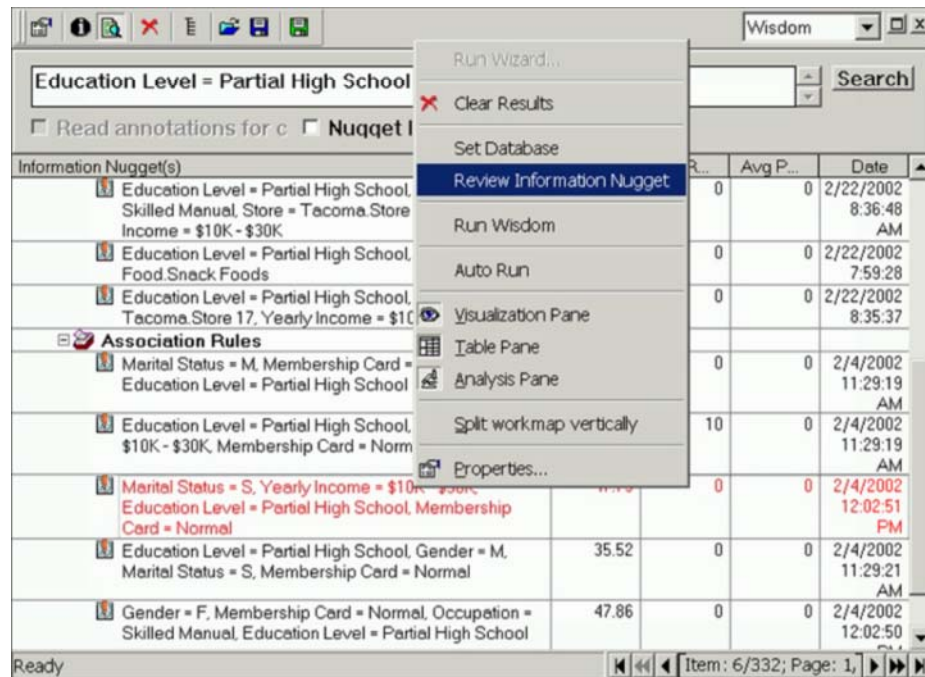
☐ Ignore Profile for current session

Reset All

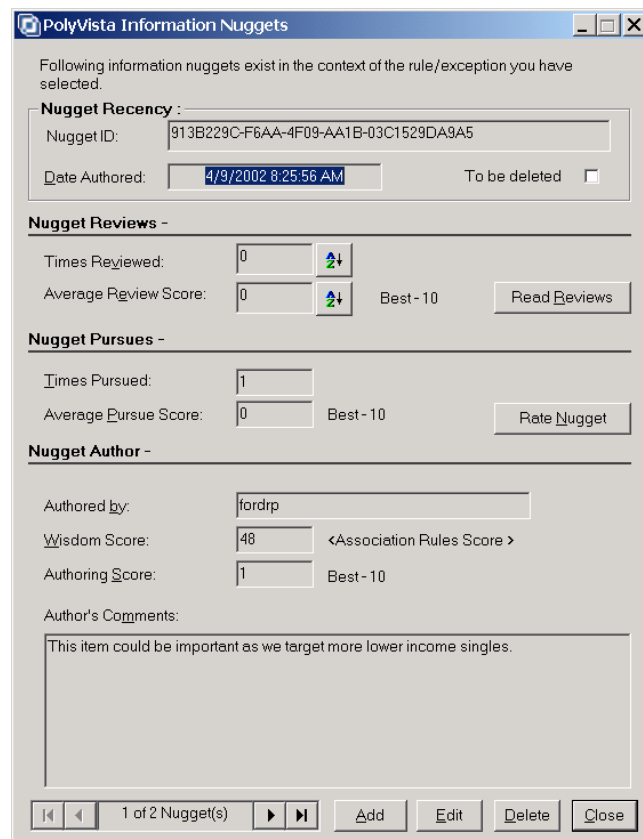
Save Close Cancel

These parameters are used to fine-tune the quality and content of the mined results list. It is used in conjunction with the users keyword list to extract only user relevant results from the Wisdom repository. Parameters can be associated with the quality of the nuggets, their recency, as well as who created them and/or who has reviewed them.

The collaborative aspects of the Wisdom repository are very important to the overall quality of a result and hence it's potential value to the business. When a user adds a nugget to the repository, they would generally add some textual context about the significance of this particular result. This information or nugget context is available to any user who subsequently finds this nugget during the search process. The next screen image depicts a search result list and the action of right-clicking the nugget of interest.



Notice the selection of **Review Information Nugget** in the previous image. Selecting this option will result in the following dialog box



The Information Nugget dialog box provides the user with additional information relevant to this particular result. Not only is the nugget author identified with their original commentary but we can also determine how many other individuals have reviewed/rated this result and we can **Read** their individual comments as well. In addition the current user can now add their comments and/or ratings to this nugget and thus add to the collective understanding (Wisdom) and business potential of this result.

In summary, these examples were presented to better illustrate general analytical processes and the overall value proposition of the PolyVista Enterprise Analytics solution. Whether the analytical requirement involves manual slice & dice methods or an automated Discovery process, PolyVista uniquely delivers results to individual users as well as the analytical community at large. By providing collaborative support to share results, interpretation and techniques among analysts and knowledge workers, PolyVista can enable and enhance the quality and value of decision support processes across a company.

