COVERSTORY - AUTOMATED NEWS FINDING IN MARKETING

John D. Schmitz
Information Resources, 200 Fifth Avenue, Waltham, MA 02154

Gordon D. Armstrong
Ocean Spray Cranberries, Lakeville-Middleboro, MA 02349

John D. C. Little
M.I.T., Cambridge, MA 02139

COVERSTORY - AUTOMATED NEWS FINDING IN MARKETING

John D. Schmitz
Information Resources, 200 Fifth Avenue, Waltham, MA 02154

Gordon D. Armstrong
Ocean Spray Cranberries, Lakeville-Middleboro, MA 02349

John D. C. Little
M.I.T., Cambridge, MA 02139

May 1990

WP#: 3176-90-MS
MCWP#: 90-5

Paper presented at

DSS-90
The Tenth International Conference of Decision Support Systems
Cambridge, Massachusetts
May 21-23, 1990

Sponsored by the College on Information Systems
The Institute of Management Sciences

Abstract

Ocean Spray Cranberries, a $1 billion fruit processing cooperative, tracks sales and assesses the effectiveness of its marketing programs with large databases collected through bar-code scanners in supermarkets. The level of detail of these databases is so great that it can easily obscure important facts and trends in the market. Ocean Spray, like many grocery companies, has a very lean staff for analyzing this data and so saw a need to automate as much of the analysis as possible. Working with Ocean Spray, Information Resources has developed an expert system, CoverStory, to uncover the important facts and trends in a scanner database. Using a series of sub-analyses, CoverStory reveals brands and markets which are currently newsworthy and identifies marketing factors that may be causing changes in these brands and markets. CoverStory results are published in an attractive memorandum format. The whole system is organized very much in the spirit of decision support. The user can adjust weights and importance criteria and can edit the final memo. CoverStory is part of an overall DSS at Ocean Spray. The scope and comprehensiveness of the system have made it possible for a single marketing professional to manage the process of alerting all Ocean Spray marketing and sales managers to key problems and opportunities and providing them with daily problem-solving information.

1. Introduction

Machine-readable bar codes on products in supermarkets have changed forever the way the packaged-goods industry tracks its sales and understands how its markets work. Although the codes were originally introduced and justified to save labor at check-out, the spin-off data produced by them provides marvelous opportunities for retailers and manufacturers to measure the effectiveness of their marketing programs and create greater efficiencies in their merchandising and promotion. We shall describe how one manufacturer, Ocean Spray Cranberries, Inc., has responded to these opportunities with an innovative decision support system designed to serve marketing and sales management.

Ocean Spray. Ocean Spray Cranberries, Inc. is a grower-owned agricultural cooperative headquartered in Lakeville-Middleboro, Massachusetts with about nine hundred members. It produces and distributes a line of high quality juices and juice drinks with heavy emphasis on cranberry drinks but also with strong lines in grapefruit and tropical drinks. The company also has a significant business in cranberry sauces and fresh cranberries. About 80% of Ocean Spray products sell through supermarkets and other retail stores with lesser amounts flowing through food service and ingredient product channels. Ocean Spray is a Fortune 500 company with sales approaching $ 1 billion per year.

Until the mid 1980's, Ocean Spray, like most grocery manufacturers, tracked the sales and share of its products with syndicated warehouse withdrawal and retail store data provided by companies such as SAMI and A.C.Nielsen. This data supplemented the companies' own shipments data by providing information on competitive products and the total market. Such databases have formed the cornerstones of useful and effective decision support systems in many companies (Little, 1979; McCann, 1988). For some time, however, it has been apparent that a radically new generation is on the way. By the mid eighties, the penetration of scanners in supermarkets had reached a level such that data suppliers could put together valid national samples of scanning stores and provide much more detailed and comprehensive sales tracking services than previously. In 1987, Ocean Spray contracted for InfoScan data for the juice category from Information Resources, Inc. (IRI) of Chicago.

InfoScan. IRI's InfoScan is a national and local market tracking service for the consumer packaged-goods industry. InfoScan follows consumer purchases of products at the individual item level as identified by the industry's Universal Product Code (UPC). IRI buys data from a nationally representative sample of over 2500 scanner-equipped stores covering major metropolitan markets and many smaller cities. These provide basic volume, market share, distribution and price information. Added to this are measures of merchandising and promotion collected in the stores and markets. These include retailer advertising in newspapers and flyers, in-store displays, and coupons. Most of the measures contain several levels of coding; for example, newspaper ads are coded A, B, or C, according to their prominence. In addition, the InfoScan service provides access to IRI's individual household purchase data collected from approximately 70,000 households across 27 market areas.

Data Explosion. The amount of data is almost overwhelming. IRI adds about 2 gigabytes per week to its master database in Chicago. Compared to the old tracking data, a company buying the InfoScan service receives increased detail by a factor of 4 to 6 because of dealing with

individual weeks instead of multi-week totals, 3 to 5 because of UPC's instead of aggregate brands, 4 to 5 because of 50 individual markets instead of broad geographic regions, 2 to 3 because of more tracking measures, and 1 to 3 because of breakouts to individual chains within a market. Multiplying out the factors reveals that 100 to 1000 times as much data is being handled as previously.

Most packaged-goods manufacturers did not initially understand the implications of two to three orders of magnitude more data. And, in fact, this kind of change is difficult to comprehend. In terms of a management report, it means that, if a report took an hour to look through before, the corresponding report with all the possible new breakouts would take 100 hours to look through. In other words, the new detail will not be looked at.

The remarkable advances that have taken place in computing have helped conceal this issue. Today's technology certainly makes it feasible to store and retrieve all the new data and, although the hardware and software to do this are not cheap, they represent a small fraction of the sales dollars involved, so that, if using the data can lead to more effective marketing, a full scale DSS with on-line access to the database is certainly warranted. Indeed, it was clear in advance and even more clear after the fact that the detailed data contain much information of competitive value in running the businesses.

DSS strategy. Packaged goods companies today have lean staffs. Many have been restructured and lost people. This is in the face of the huge data increases just described. Although Ocean Spray has not been restructured, its roots as an agricultural cooperative have always given it an internal culture of lean self-sufficiency. It has a small IS department for the organization as a whole.

This situation led Ocean Spray naturally to a strategy of a small marketing DSS organization running a decentralized system where the users do most of their own retrieval and analysis. The marketing DSS for syndicated data currently consists of one marketing professional plus the database administrators. The goal is to have a largely centralized database with workstations for sales and marketing in the business units. User-interfaces must be easily mastered by busy people whose main jobs are in the functional areas. The role of the DSS organization is to acquire and develop tools with which the end-users can do their own analyses. DSS consults with users to develop appropriate pre-programmed reports to be delivered as hard copy and/or on-line.

An important characteristic of the system must be growth potential. Not only should retrieval of specific numbers, tables, and graphs be easy now, but the system architecture and computing power should be there for future calculations and analyses that are likely to be much more computationally intensive than simple retrieval.

Ocean Spray's InfoScan database. Ocean Spray's syndicated database for juices is impressive, almost imposing, considering the change from the past and the level of human resources put against it. It contains about 400 million numbers covering up to 100 data measures, 10,000 products, 125 weeks, and 50 geographic markets. It grows by 10 million new numbers every four weeks. Finding the important news amid this detail and getting it to the right people in a timely fashion is a big task for a department of one.

Hardware and Software. The DSS architecture puts the database and CPU-intensive

processing on an IBM 9370 mainframe with ten gigabytes of disk storage and puts user-interface tasks on eleven 386-level workstations located in the marketing and sales areas. The basic DSS software is IRI's DataServer, which manages data and mainframe computation in the fourth generation language EXPRESS and the user interface in pcEXPRESS. This provides menu driven access to a family of flexible, pre-programmed reports available on the workstations.

Unlike some other solutions used by packaged-goods manufacturers, this architecture provides easy access to mainframe computing power from the workstations as is needed, for example, to run applications like the CoverStory software to be discussed.

2. Basic Retrieval and Reporting

The basic retrieval, reporting and analytic capabilities of Ocean Spray's DSS are extensive. Any particular fact from the database can be pulled out in a few steps with the help of pull-down menus and picklists. Much of the use comes from standard reports: a company top line report, and four business area reports (cranberry drinks, grapefruit, aseptic packages, and tropical drinks) showing status and trends including changes in share in aggregate and in detail, and changes in merchandising and distribution against a year ago or four weeks ago. Derived measures such as BDI (brand development indices) and CDI (category development indices) are available. Product managers can get a quick update of what is going on with their products. Standardized graphs can be called up and it is relatively easy to construct new ones. Similarly, users can readily construct measures that are ratios, differences, and other combinations of ones already in the database. Usage has been growing steadily since DataServer and the InfoScan database were been installed.

Nevertheless, the introduction of the system has required as much learning for the DSS department as the end users. Some people, especially in sales, made little use of the system. Within marketing a few individuals took to the system quickly and did considerable analysis but there was also a feeling that you would not want to have a reputation for spending too much time pushing numbers around. In fact, within sales, the characteristic attitude has been: "Using the computer is not my job. Give me something that is already analyzed. Give me materials that are ready to use and will help me do my job."

In response to this the DSS department has developed (and continues to develop) tools and analyses that will help solve specific user problems. There are a number of approaches; CoverStory to be discussed below is one of the key directions. In addition a variety of reports oriented about selected issues have been developed. For example, reports that rank products and point out Ocean Spray strengths and identify markets where some Ocean Spray product is underdistributed relative to its inherent selling power. The intention is to let sales and marketing people identify market opportunities and product selling points.

3. Finding the News: CoverStory

CoverStory is an expert system developed by IRI to tackle the problem of too much data; Ocean Spray has been a development partner and first client. CoverStory automates the creation of summary memoranda for reports extracted from large scanner databases. The goal is to provide a cover memo, like the one a marketing analyst would write, to describe key events that are

reflected in the database - especially in its newest numbers. The project began as a teaching exercise in marketing science - "How would you summarize what is important in this data?" (Stoyiannidis, 1987; Little, 1988) - and has developed into a practical tool.

CoverStory is undergoing continuing development as we gain experience with its use in new situations. We describe the following aspects of the system as it now now being used: (1) the role of marketing models, (2) the basic decomposition steps embodied in the search strategy, (3) the linearization and ranking processes used to decide what facts are most worth mentioning, and (4) methods for generating and publishing the output.

Marketing Models. CoverStory is rooted in the modeling tradition. Howevr, by design, it does not directly present model results at this stage of development, but rather reports only database facts, such as share, volume, price, distribution, and measures of merchandising. The reason is to have the output and underlying processes as transparent and easy to understand as possible. The program assesses the relative importance of these facts, and selects them for presentation by using weights and thresholds which come from marketing models. However, the user is able to inspect and change these values.

Futhermore, in choosing measures of marketing effort for CoverStory to consider, we select a set of marketing variables from the scanner database that model-building experience has shown to be important for driving sales and share. Measures commonly used include:

Displays - percent of stores (weighted by size) that displayed a brand or item.

Features - percent of stores (weighted by size) that ran a feature ad on a brand.

Distribution - size-weighted percent of stores that sold a brand.

Price cuts - percent of stores that sold a brand at a price reduced by more than a threshold, such as 10%, from the regular price.

Price - Price can be represented by many data measures. CoverStory sometimes uses the overall average price paid at the register but often draws from a finer set of price measures that may include regular price, average merchandised price, and depth of discount. The regular price is the price of an item not undergoing special promotion; average merchandised price is the price of an item in stores where it is being promoted with feature ads or displays. Depth of discount is the difference between these two. In an InfoScan database, we can get even finer measures of price by breaking out average merchandised price and depth of price discount by type of merchandising.

Through a marketing model, we quantify the impact of each of these marketing levers on share or on sales volume and find their relative importance. For grocery items, among the measures described above, we usually (but not always) find that distribution is most important then price then displays then features then price cuts.

Flow of Analysis and Decomposition. Figure 1 shows the general flow of analysis in CoverStory. The central idea is that we will analyze the behavior of an aggregate product in an

aggregate market by a series of decompositions or disaggregations. An aggregate product is a product which includes more than one UPC. The UPC (bar code on the package) is the lowest level of product detail available in a scanner database. An example of an aggregate product is Ocean Spray Cranberry Juice and Blends. It consists of many different sizes, package types, and flavors and blends. An example of an aggregate market is the total U.S. which can be disaggregated into regions or individual cities or even grocery chains within markets.

In doing decomposition, CoverStory follows a style which we have observed in the analytical marketing reports used in many companies. Analysis proceeds by answering the following series of questions. (1) What is going on overall in the aggregate product for the aggregate market? (2) What changes does this reflect in the components of the aggregate product? (3) What changes does this reflect in the components of the aggregate market? (4) What is happening to competitive products? In CoverStory, we go through each of these in turn. Within each of these sections of the analysis, the program follows a standard series of steps:

> Rank the components (markets or products or market/product combinations) by some criteria.
>
> Select the most noteworthy for mention and for further analysis.
>
> Calculate causal factor changes for these top few markets, products, or combinations. Causal factor changes are distribution, price, and merchandising changes.
>
> Rank these causal factor changes then select the top few causal changes to include in the report.

The need to select "the top few" items from different lists is dictated by the size of the scanner database. The number of events that can be mentioned is enormous. Without strictly limiting the amount of information in a report, we found that the news drowned in the detail.

Ranking the products or markets. We nearly always rank component products or component markets by share or volume change. When we are looking at size groups within an aggregate product, for example, and we are analyzing share changes, size group ranks will be based on share changes. We have found this to be generally effective with one exception. If there has been a fundamental restructuring of the way a category is marketed, share changes may not be meaningful. This showed up when we looked at coffee sales. In late 1987 and early 1988, coffee packaging switched from packages that were multiples of a pound (1, 2, and 3 pound cans) to packages that were multiples of 13 ounces (13, 26, and 39 ounces). This gave the appearance that a large amount of volume was switching into "new products" and, for a year, volume and share change calculations required special treatment.

Selecting the top few products or markets. The top few are the few that are the most noteworthy. We generally calculate which component products or markets are furthest away from average and retain these extremes for mention. Normally, this leads CoverStory to pick winners and losers. In some cases, however, when most of the products and markets are behaving in similar

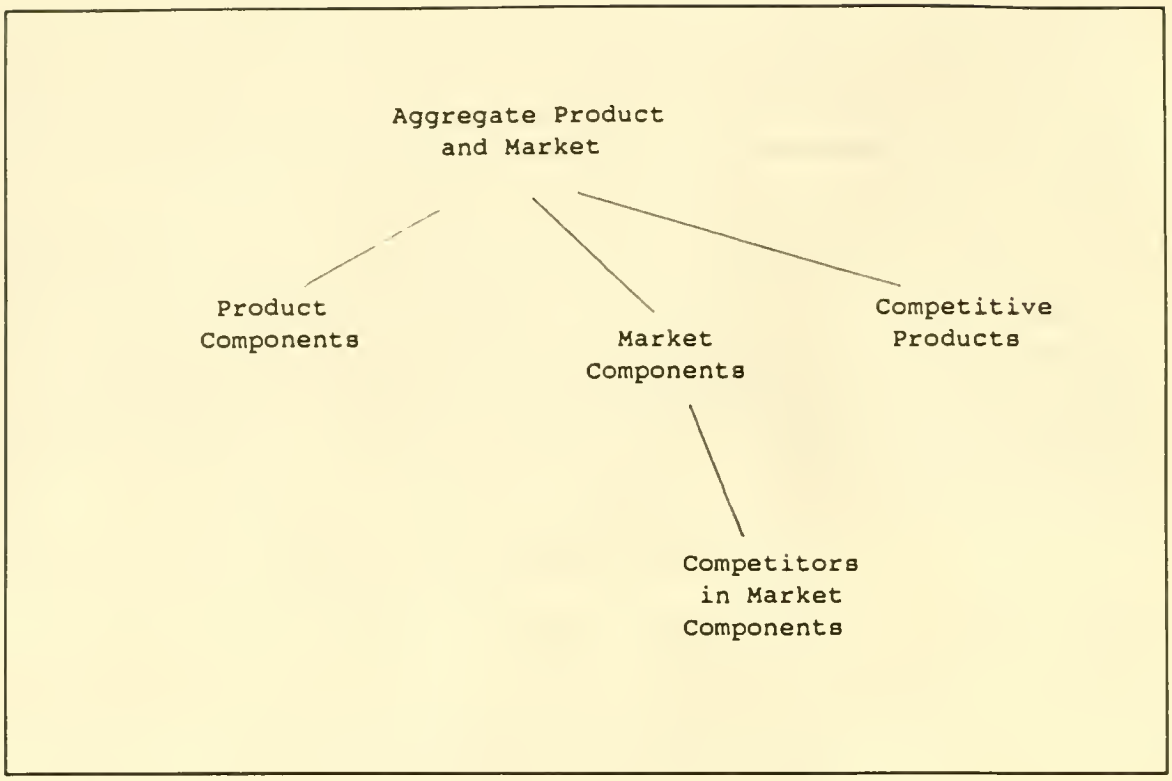


Figure 1 – CoverStory Analysis

fashion, CoverStory will select only winners or only losers. This approach has been very effective and it closely mimics the way that human market analysts select individual segments of a product line or individual markets for mention.

Calculating causal factor changes. When we point out share or volume changes, we would like also to mention possible causes of these share or volume changes. To do so, we calculate the amount of change in marketing support in each of the marketing factors which affects the product. For example, a share change in CranApple sales in Boston may have been partially caused by distribution, price, display, feature, or price cut activity.

Ranking Causal Factor Changes. We can generate a large number of causal factor changes when we decompose the aggregate product and market behavior into components. If there are ten product components, fifty markets, and eight causal factors we are screening, we have four thousand causal changes which are candidates for mention. Trimming this down to a small number for inclusion in the CoverStory report requires a ranking procedure. The procedure we have chosen is similar in spirit to the evaluation functions used in evaluating positions in game-playing programs (Barr, 1981). We calculate a score for each of the causal measure changes. The score incorporates the market in which the change occurred, which causal factor changed, and the magnitude of the change. Symbolically:

Score = Change * Factor weight * Market weight

Change is the amount of change in the causal factor and is either a percent change or raw

change depending on the factor. Factor weight is different for each of the marketing factors such as distribution, price, displays, featuring, and price cuts. These factor weights are intended, informally speaking, to make different marketing changes have the same score if their impact on sales is the same. We initialize factor weights based on analysis done outside of CoverStory based on logit models of the type described in Guadagni and Little (1983). Market weight is a term which makes it more likely that an event in a large market will be mentioned than an event in a small market. We originally used market size but found that this was too strong. Only events from New York, Chicago, and Los Angeles would be mentioned and so we have softened the impact of market size. One approach that has proven effective is to use the square root of market size as the market weight.

In all, this scoring method yields a ranked list of causal market changes where such a change can be described in terms of:

What happened? (e.g. price went up by 20%)
Where did it happen? (e.g. in the Southeastern Region)
What product did it happen to? (e.g. the 32 ounce bottle)

The events that CoverStory describes are the ones that rank highest using this scoring mechanism.

Presenting the Results. We have experimented with several methods for presenting these results. Our present style is to produce an English language report in distribution-quality format. This has been an important piece of the overall effort and has had a dramatic effect on the acceptability of CoverStory reports to end users. The language generation is usually straightforward; it is based on sentence templates (Barr, 1981). We have considered but not yet implemented context and memory (Schank and Riesbeck, 1981) in our text generation. The use of some randomization of detailed wording through the use of a thesaurus keeps the CoverStory memo from sounding too mechanical. The memo is relatively short and structured so this simple language generation has not been a limitation on CoverStory.

The CoverStory results are published through a high-quality desktop publishing package or a word-processor with desktop publishing capabilities. Variation in type-face, use of graphic boxes, and sidebars are all intended to give the memo visual appeal and highlight the marketing facts which are contained in it.

CoverStory is very much a decision support system rather than a decision making system. The user can adjust all major system parameters such as who competes with whom, what weights to use for the marketing factors, and how much information is to be reported. The final memo is published through a standard word processing package so it can be edited by the user, although this seldom happens. Because the memo is automated and easily set up (and then left alone) to meet the needs of specific managers, the appropriate "news" can quickly be distributed throughout the organizations when new data arrives.

A CoverStory memorandum attached as Figure 2 illustrates the output. In this coded example, we present highlights about a brand called Sizzle in the Total United States. The recipient for this memorandum is the Sizzle Brand Manager and the brand management team. The series of decompositions in this report is:

To: Sizzle Brand Manager

From: CoverStory

Date: 07/05/89

Subject: Sizzle Brand Summary for Twelve Weeks Ending May 21, 1989

Sizzle's share of type in Total United States was 71.3 in the C&B Juice/Drink category for the twelve weeks ending 5/21/89. This is an increase of 1.2 points from a year earlier but down .5 from last period. This reflects volume sales of 10.6 million 192oz equiv. Category volume (currently 99.9 million 192oz equiv) declined 1.3% from a year earlier.

Display activity and unsupported price cuts rose over the past year - unsupported price cuts from 38 points to 46. Featuring and price remained at about the same level as a year earlier.

Components of Sizzle Share

Among components of Sizzle, the principal gainer is:

Sizzle 64oz: up 2.2 points from last year to 23.7

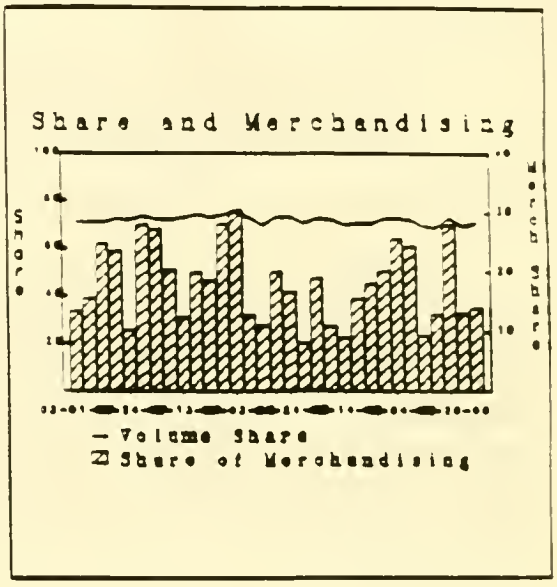
and losers:

Sizzle 48oz -0.6 to 34.9

Sizzle 32oz -0.1 to 6.7

Sizzle's share of type is 71.3 - up 1.2 from the same period last year.

Sizzle 64oz's share of type increase is partly due to 11.3 pts rise in % ACV with Display vs yr ago.



Competitor Summary

Among Sizzle's major competitors, the principal gainers are:

Shakey: up 2.5 points from last year to 2.6

Private Label +.5 to 19.9 (but down .3 since last period)

and loser

Generic Seltzer -.7 to 3.5

Shakey's share of type increase is associated with 71.7 pts of ACV rise in

Breaking down total Sizzle volume into sales by size groups.
Looking at Sizzle's major competitors.
Looking at submarkets of the US - cities in this database.
Looking at competitive activity in these submarkets.

The analysis is based on share change. A sample of a causal change shown by CoverStory is the increase of display activity to support 64oz bottles of Sizzle.

4. Benefits

Ocean Spray's DSS design strategy has successfully solved several problems. The decision to put users in charge of their own basic retrieval and analysis has generally worked well and, where it has run into problems, the DSS organization has responded by providing increasingly customized tools. The DataServer interface has been easy to learn. Usage on the 386-level workstations located in the marketing area is many hours per week and rising.

The strategy casts the DSS organization in the role of acquiring and building tools to make the users more effective. Consultation with users has led to a set of hard copy reports that are circulated regularly to marketing, sales and top management and to customized reports that can be called up on line and printed locally on laser printers, if needed.

CoverStory is a particularly desirable development because, with very little effort, it provides users with top line summaries and analyses across a wide variety of situations. Previously this required time-consuming intervention by a skilled analyst. Furthermore the technology is an extensible platform on which to build increasingly sophisticated decentralized analysis for the user community.

The information coming out of Ocean Spray's marketing DSS is used every day in planning, fire-fighting, and updating people's mental models of what is going on in the company's markets. Typical applications include such actions as taking a price increase and monitoring its effect; discovering sales softness in a particular market, diagnosing its causes, and applying remedies; and following a new product introduction to alert the sales department in case of weak results in certain markets compared to others. The DSS is totally integrated into business operations and it is no longer seems possible to consider life without it.

Perhaps the easiest way to express the success of the system is that, with the help of marketing science and expert systems technology, the DSS has made it possible for a single marketing professional to manage the process of alerting all Ocean Spray marketing and sales managers to key problems and opportunities and of providing them with daily problem-solving information and guidance. This is being done across four business units handling scores of company products in dozens of markets representing hundreds of millions of dollars of sales.

REFERENCES

Barr, Avron & Edward A. Feigenbaum eds., Handbook of Artificial Intelligence, Los Altos, CA: William Kaufman Inc. 1981.

Guadagni, Peter M. & John D.C. Little, " A Logit Model of Brand Choice Calibrated on Scanner Data," Marketing Science, 2 (Summer 1983), 203-38.

Little, John D.C., "Decision Support Systems for Marketing Managers," Journal of Marketing, 43 (Summer 1979) 9-26.

Little, John D.C., "CoverStory: An Expert System to Find the News in Scanner Data," internal working paper, Sloan School of Management, M.I.T. (September 1988).

McCann, John M., The Marketing Workbench, Dow Jones-Irwin, Homewood, IL, 1986.

Schank, Roger C. & Christopher K. Riesbeck eds., Inside Computer Understanding, Hillsdale, NJ: Lawrence Erlbaum Assoc. 1981.

Stoyiannidis, Demosthenes, "A Marketing Research Expert System," Sloan School Master's Thesis, M.I.T., Cambridge MA 02139, June 1987.

1-8-91