Money, the Internet, and Jai-Alai

Steven Skiena

Department of Computer Science
State University of New York
Stony Brook, NY 11794–4400
http://www.cs.sunysb.edu/∼skiena

O'Reilly Money:Tech Conference
February 7, 2008

Program Trading in Jai-Alai

Investing in the financial markets is often likened to gambling.

This is story of how I used computer simulations and mathematical modeling techniques to bet on jai-alai.

The Power of First Principle Modeling

I tell the “Calculated Bets” story to emphasize the power of first principle modeling; explaining why things happen as opposed to detecting things that happen to happen.

Early History

I first learned to program a computer as a high school student back in 1977, when I developed a simple little model to predict the results of NFL football games.

This was a big enough deal back then for me to become a newspaper columnist. . .

Student uses computers to predict football winners

By JEFF LEEBAW
Home News staff writer

EAST BRUNSWICK — A 16-year-old East Brunswick High School student has found a way to combine an interest in football with a fascination for computers.

Steven Skiena says he can determine, with a high degree of accuracy, the outcome of professional football games by feeding a computer pertinent information about competing teams.

"The winners will almost always be correct," said the high school junior who lives at 5 Currier Road off Dunhams Corner Road. "I had an 86 per cent accuracy rate when I started predicting at the end of last season."

He does it by feeding the computer a myriad of statistics that include team records, points scored and allowed, average yards gained and allowed during a game, a breakdown of the yards gained and allowed into rushing and passing categories, performances at home and on the road, and more.

The information is gathered from weekly compilations of football statistics and standings. Skiena puts the facts on index cards and then types them into one of the six computer terminals at the high school or a terminal at The Library where he works part-time after school.

"I get a winning team, a decimal score for each team and a point spread," said the teen-ager who completed a computer programming course last year at the high school.

His first attempt at picking winners involved a Monday night game between the Oakland Raiders, the eventual Super Bowl victors, and the Cincinnati Bengals.

It was a difficult game to analyze because Cincinnati was fighting for a playoff berth while Oakland had already clinched a spot in the post season competition.

"Nobody knew whether Oakland would be giving 100 per cent," Skiena said. "But my calculations indicated they would win by 24-20. The final score was 35-20. They went all out."

Skiena said he went on to pick 12 of the 14 winners the following week and accurately predicted Oakland would defeat Minnesota in the Super Bowl.

The National Football League's 1976 Record Book, which breaks down last year's statistics for each of the league's 28 teams, will supply most of Skiena's information for the first few weeks of the 1977 season. He will also use statistics from the final two exhibition games played this year by each of the teams.

Skiena wrote a computer program based on 17 statistical variables that might come into play during a football game.

STEVEN SKIENA
...to test his accuracy

Predictions published

Steven Skiena will get a chance to display his skill as a pro football prognosticator each Sunday in The Home News.

The youngster's weekly selections will be an "added ingredient to our football coverage," according to Home News Executive Editor Robert E. Rhodes.

"I think it's interesting enough for us to give it a

Lessons from Football Prediction

1. Simple mathematical models have real predictive power
2. Simple mathematical models do not have predictive power when money is on the line

I also got interested in jai-alai during high school, after being exposed to the game on a family vacation to Florida.

What is Jai-Alai?

Jai-alai is a sport of Basque origin also known as *Pelota Vasca* or *Cesta Punta*.

It is a variation on hand ball, where the rubber-and-goatskin *pelota* is caught/thrown with an large basket or *cesta*, and bounced off a granite or concrete wall.

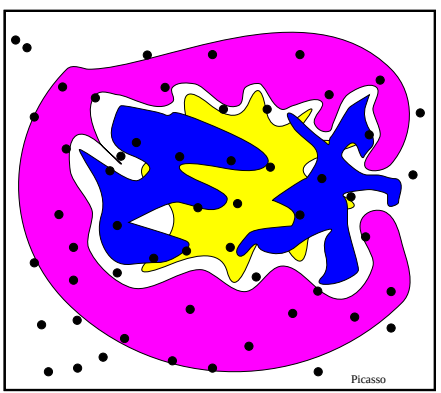
Why Jai-Alai?

The scoring system in use in the United States is different than that of the Basques, and particularly interesting to a mathematician or computer scientist.

The *Spectacular Seven* scoring system involves a queuing system that produces very biased results depending upon where you start in the queue.

Monte Carlo Simulation

Monte Carlo techniques use many random experiments to approximately measure a real quantity, as in Monte Carlo integration.



The biases inherent in the Spectacular Seven Scoring System can be identified by simulating each match 1,000,000 times.

Trifecta Results

Certain trifectas occur 1,000 times more often than others, assuming equally skilled players!

Several trifectas are *unbelievably* terrible bets, such as 5-7-8, 5-8-7, 6-7-8, and 6-8-7. In 29,096 games at Dania jai-alai, these four trifectas (and only these four) never happened.

Our system mostly bets trifectas.

The greatest inefficiencies in any market reside in the most complex derivatives.

Getting and Parsing Data

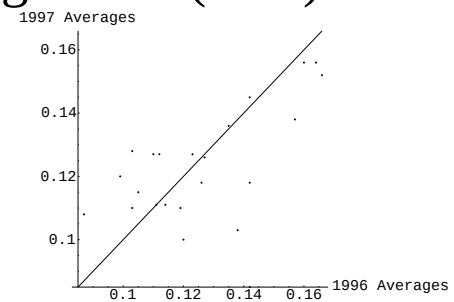
These results are not enough to bet successfully, because such biases are basically captured in the daily odds.

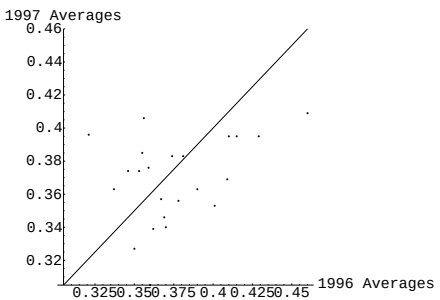
We need to analyze statistical data to measure player skills and estimate payoffs.

Each night our spider programs got the latest data from the WWW, parsed it using Perl, and added it to our database.

Are There Good Players and Bad Players?

Correlations in year-to-year performance of jai-alai players (0.49) is comparable to baseball batting averages (0.52) and pitching ERAs (0.50).





Thus we *can* use history to evaluate player skills in a meaningful way.

Variance in returns is analogous to market risk.

How Much Does Each Bet Pay?

Jai-alai frontons make their money by taking a cut of all bets, under the *parimutuel* system.

Frontons want you to bet, casinos want you to lose. Thus they were happy to take my bets even if my system wins.

By averaging the results of previous payoffs, we can estimate the likely payoff for every possible bet.

Brokers want you to bet, too.

How Much Do We Impact the Pool?

Our successful bets will lower the payoffs on these outcomes, eventually leading to negative expected return.

This is why real investors bet at the stock market instead of the fronton.

Jai-alai pools are small enough that very modest investments flood the market, but we can predict the impact of our own wagering.

Managing market impact is a significant part of controlling transaction costs.

Putting the System Together

Each night, our system downloaded the schedule of tomorow's matches, predicted the results, and identified the best bests.

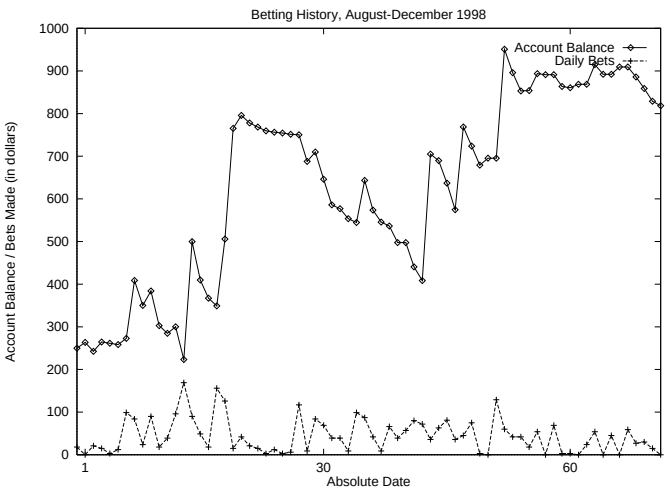
In the morning, our computer phoned the bets in (using a modem) to the fronton's computer.

There was no human involvement in analysis or wagering!

Order execution is an important component of serious investment strategies.

How Did We Do?

We carefully logged our bets from July to December, 1998.



This is a total capital return of 227% over the period, with a return of 18% per bet.

Epilogue: Jai-alai

A revised system later increased our total return to 574%.
Unfortunately, Milford Jai-alai closed in December 2001.
All winnings were donated to charity.
But what's next after modeling Jai-alai. . .

Modeling the World

We now want to model much larger parts of the political, social, and economic world.

This requires knowledge of how all the world's major entities (people, places, companies, products, commodities) interact with each other.

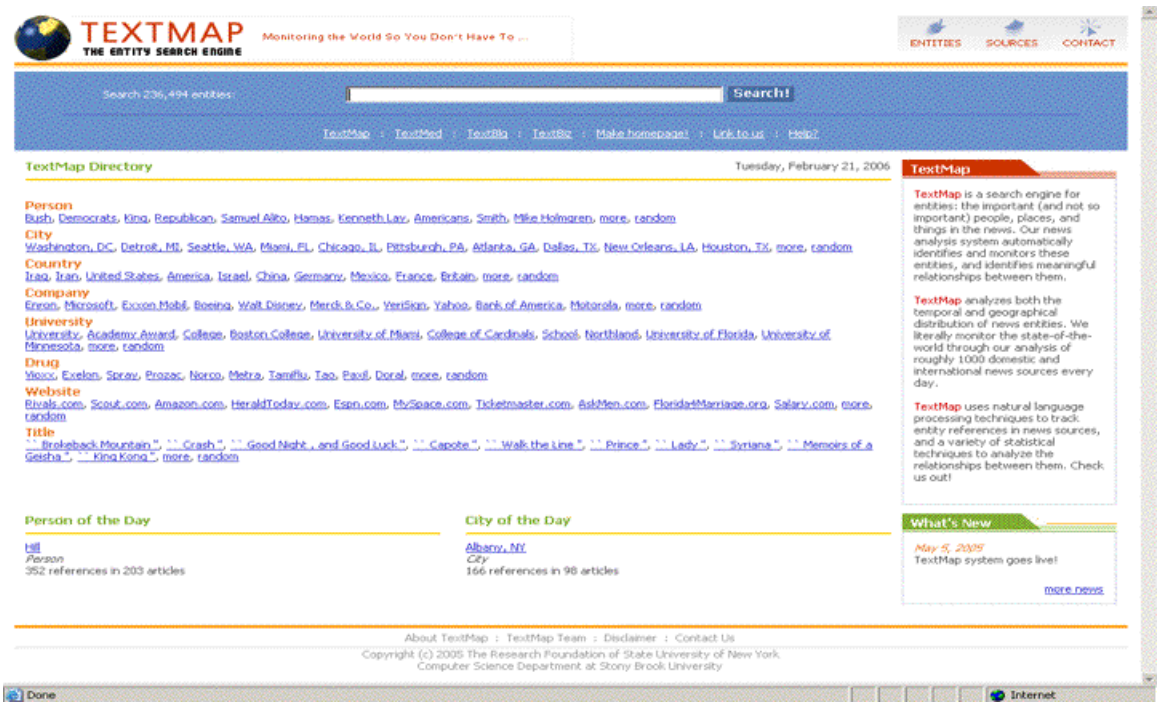
Obtaining this knowledge requires extensive computational analysis from news, blogs, and other text streams.

Lydia News Analysis

Our *Lydia* news analysis system does daily natural language analysis of over 1000 English and foreign-language newspapers plus blogs, RSS feeds, and other text streams.

We currently track over one million news entities, providing spatial, temporal, relational, and sentiment analysis.

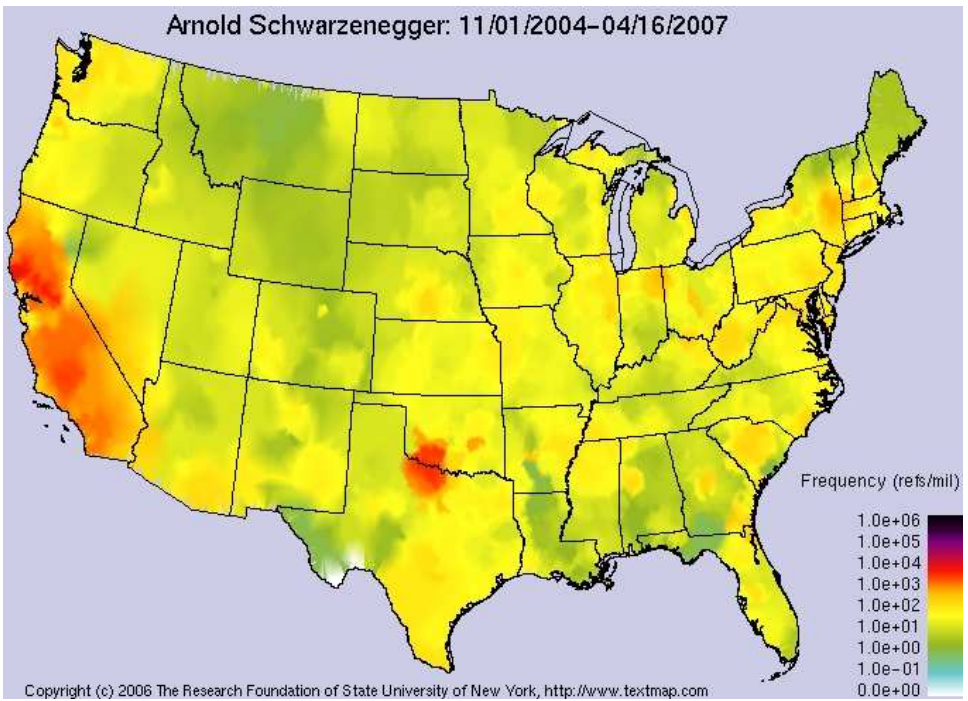
www.textmap.com



Lydia System Architecture

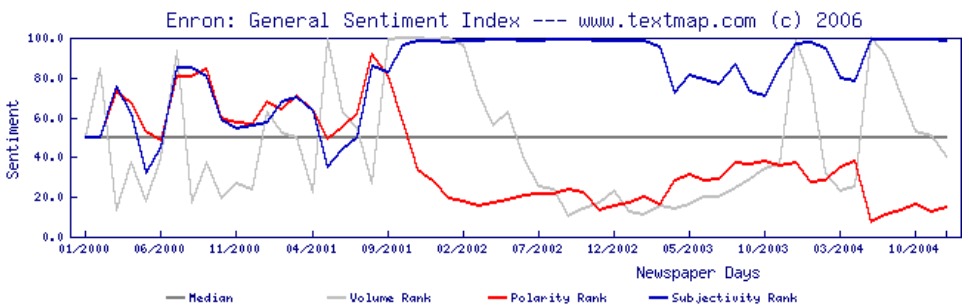
- Spidering and Normalization – Text is retrieved using semi-custom spidering agents and parsers to format for our pipeline.
- Text Markup – Annotates interesting parts of the source text using natural language processing (NLP) techniques.
- Back Office Operations – We aggregate entity frequency and relational data into a MySQL database for statistical analysis.

Spatial News Analysis: Reference Frequency



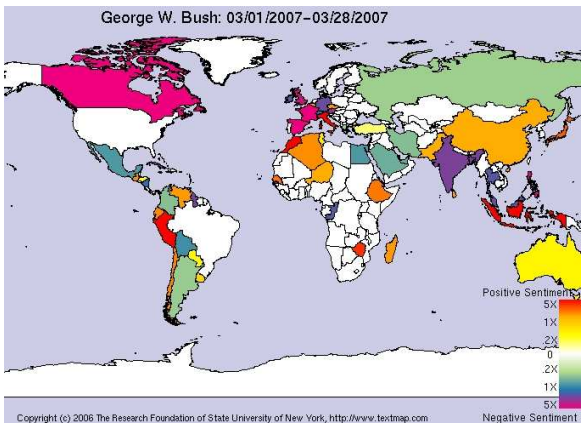
Sentiment Analysis: When Did Enron Go Bad?

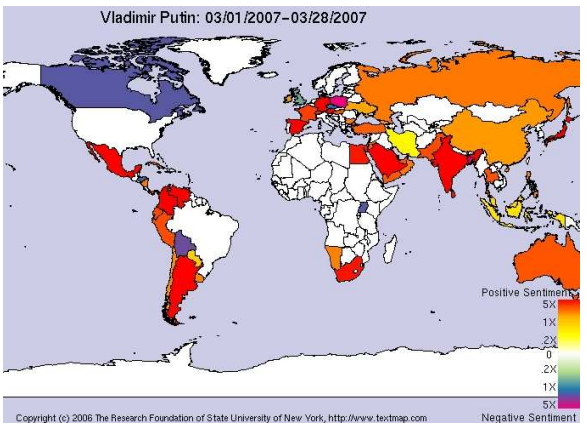
Sentiment analysis lets us measure how positively/negatively an entity is regarded, not just how much it is talked about.



Detecting sentiment requires sophisticated analytical methods.

Bush/Putin: Who is More Popular Where?

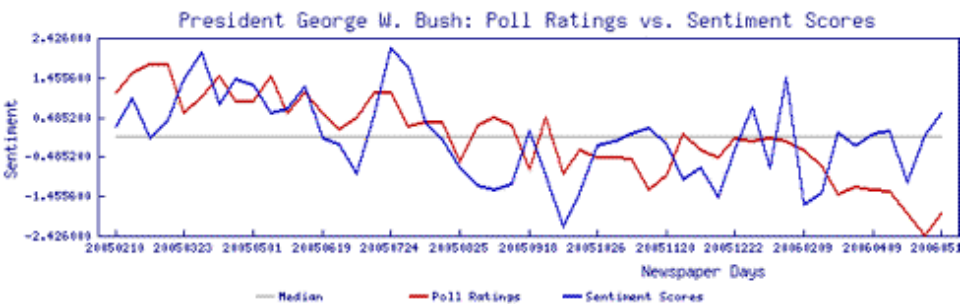




Our methods work even in foreign-language news sources.

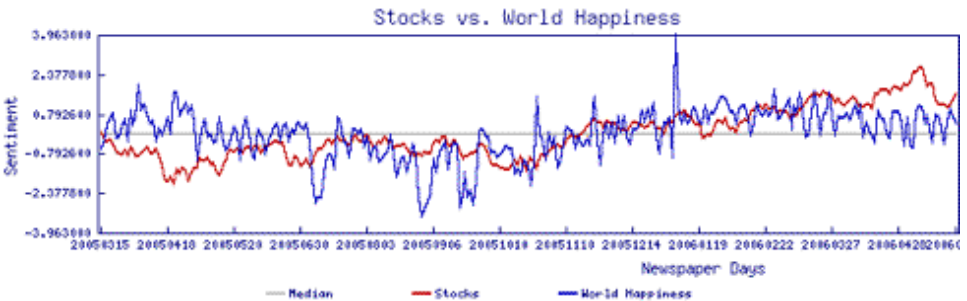
What Does it Mean?: Public Sentiment

Our sentiment scores from news analysis agree very well with traditional (but expensive) polling methods.



What Does it Mean?: Market Sentiment

Daily news sentiment accurately correlates with the state of the financial markets:

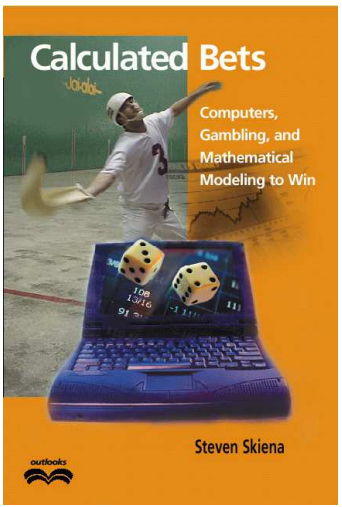


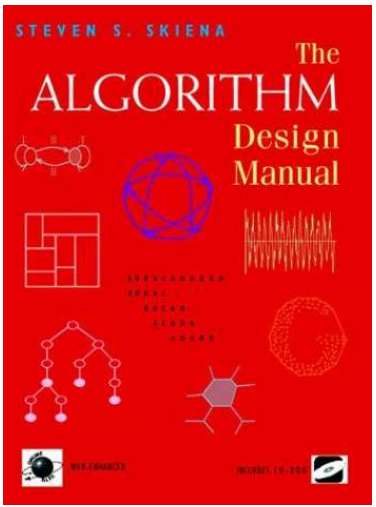
Current Directions

- Foreign-language news analysis
- Financial modeling and analysis
- Market research analysis
- Entity-oriented search engines

We actively seek partners interested in our technology.

For Further Reading – www.jai-tech.com





Don’t forget to check out www.textmap.com!