COLLABORATIVE FILTERS
by Jerry Michalski

Can’t decide which CD to buy at the local Virgin Megastore? You’re not alone, and a bunch of technology startups are banking on it. They would like to get you over that hump more quickly by plying you with recommendations from soulmates in cyberspace. That’s the effect, anyway, of using collaborative filtering technologies that Firefly Online and other Websites are popularizing.

Music is in fact the most common early application for these filters, which makes sense. It is possible to play music online; a short sample is often enough to get a feel for a song or even an artist. Individual titles are relatively inexpensive and available everywhere, including online. Music has many known genres and most people are familiar with hundreds of titles.

Collaborative filtering technology seems promising in many other domains, including literature, software, Web pages, games and news; dating services; hospitality (e.g., hotels, bed-and-breakfasts, restaurants); teachers, courses and schools; investments; and health care (doctors, hospitals, fitness clubs and products, and spas).

Combined with PICS, the Platform for Internet Content Selection, filters could help communities create their own content rating systems. The basic requirements are that selections depend on subjective judgment and domains attract enough people to get statistically meaningful results.

Of course, it’s not quite that simple. Companies that sell collaborative filtering technology and services are only beginning to test their systems’ accuracy — a difficult thing to measure. Some domains may prove intractable for any number of reasons, including participants’ conflicting motivations or simply a scarcity of people interested in giving opinions.

In the most optimistic scenario, people can (and do) rate any item or service they come in contact with, and their observations feed back into a turbocharged cross between Zagat’s, the Better Business Bureau and the Usenet. Manufacturers and service providers use this information to improve their offerings — all while ostensibly protecting everyone’s privacy.

PC FORUM INVITATIONS WITH THIS ISSUE!

INSIDE
COLLABORATIVE FILTERS 1
Sharing -- sort of.
Those pesky agents!

THE FILTERS 7
Firefly Network.
NetAngels.
Surf Communications.
Net Perceptions.
Net Deva and others.

TRUST 19
Privacy; eTrust.
Closing the loop.

RESOURCES & CALENDAR 22
Names, phones, dates, URLs.
Whom should you trust?

Collaborative filtering systems can collect an amazing amount of information about users and their activities, especially when they are enhanced with client-side software that collects explicit user feedback. Participants who use the feedback mechanisms heavily in the hope of getting better recommendations from the system are also sorting themselves into categories far more specific than "suburban soccer moms."

Most systems request e-mail addresses from new registrants. They might as well ask for their true names; they're relatively easy to look up, given e-mail addresses. Even if the company running the filtering system doesn't make an effort to learn people's true names, it ends up knowing a lot of very specific user information. Combine that information with data from other activities such as purchases, and (so marketers hope) companies can put consumers squarely in their marketing crosshairs.

The intermediaries must strike a delicate balance in their use of this information. The same datum that can make life considerably easier for a customer in one place can destroy that person's privacy if it goes elsewhere. Companies are working together to establish common privacy expectations and the procedures needed to maintain it through groups such as eTrust, which is sponsored by the Electronic Frontier Foundation. We return to that topic later (see page 19).

A filtered world

This issue of Release 1.0 explores automated collaborative filtering (ACF) technology, its uses and potential misuses, with brief excursions into the related topics of agent software and privacy issues. The companies we cover include NetAngels, Firefly Networks, Surf Communications, Net Perceptions and Net Deva.

Collaborative filters, part of the larger category of recommender systems, collect user ratings of items in a domain (e.g., movies, books, people) and use them to offer recommendations of other items in the domain. The simplest approach, calculating average ratings for each item, would rapidly lose significance as people with different tastes rated the items, effectively counteracting each other's opinions. Collaborative filters create "taste spaces" within which ratings tend to cluster. By comparing rated items between neighbors in a cluster, a filter can offer personalized recommendations efficiently, without direct contact between participants.

If collaborative filtering systems prove themselves over the next year, it is likely that service providers, merchants and others will adopt the technology. For example, as AOL moves to flat pricing next month, it will have to master these techniques to make users' experiences more fulfilling and to generate new revenue streams.

Flipping the model can change its uses dramatically. Instead of pointing people to items they might find interesting, an ACF system could help messages, products and advertisements find their way to the right people.

Finally, ACF systems are likely to affect market dynamics. For example, they should help broaden the music market by offering attractive alterna-
tives to top-40 lists, which means more attention (and potentially revenues) for obscure artists.

SHARING -- SORT OF

Automated collaborative filtering is not the most intuitive or mellifluous name for a technology that helps people share their opinions, albeit indirectly. In ACF systems, direct interaction among participants is the exception, not the rule. People interact with the system in order to hear back from the system what other people have told it.

Most Websites that use collaborative filtering do go the extra step to help participants get in touch with one another, but that feature is outside the filters themselves. In fact, filtering is a feature, not a standalone service. Success for companies using ACF depends on weaving the filtering technology into a richer environment that combines many ways to interact and navigate, without confusing or overwhelming participants.

Visitors to a site that uses ACF technology are asked to rate items -- typically movies or album tracks, although early projects have also used Web pages, Usenet postings and news items as fodder. The systems then search for patterns in the ratings by applying various algorithms that are these companies’ trade secrets. Then they point participants to novel items based on what people with similar tastes indicated they liked. (For further explanation and a comparison of several systems, see pages 7 & 16.) Each time a list of items comes up, users can rate them, which should increase the systems’ accuracy.

Timing is everything

Time is important in collaborative filtering. The results any individual gets from an ACF system should improve over time, as the individuals rate more stuff themselves and as everyone else rates items, building the matrix of ratings. Participants who stay build trust in the system, but newbies face a cold-start problem: They have to invest effort in the filter to reap its benefits. The data-entry burden can be significant.

Nevertheless, successful first use is critical. It’s easy to tell whether an ACF system is working for you: Some of the recommendations it makes will be familiar. If successive recommendations don’t converge quickly toward items you like, you will lose confidence and interest in the system.

For example, our first experience with Firefly wasn’t all that satisfying. We spent several hours of patient entering and rating and were rewarded with recommendations that included a few unappetizing-sounding groups plus Led Zeppelin, which does not lie on our vector of preferences. (Unfortunately, because it is a showcase site and not a full-function service, Firefly Online doesn’t yet offer the ability to listen to samples of music or to buy albums electronically.)

Part of the problem was our bent toward Latin and world music, which Firefly members don’t seem to listen to much, though they like it when they do. (But if you like Smashing Pumpkins, Nine Inch Nails or Seal, you’re in business.) We look forward to people from other nationalities coming into the system, so Anglo US members can take advantage of Italian members’
Some context

The interest in filters is a natural step in the co-evolution of the Net and the older computing and communication worlds. In the short period since the Web caught fire, companies have already moved through several phases of Web use. They started with obvious stuff: Web front doors, service descriptions and product catalogs. Excited about the Web’s potential, many companies rushed to build or buy commerce engines, virtual stores and other transaction systems, but aside from a few notable exceptions, the timing was off. It was too early for transaction systems to generate much commerce. Not enough people were online yet; of those, too few had secure Web connections or digital payment systems (e.g., CyberCash, First Virtual). Even people who bought things were annoyed by the effort to register and fill out shipping information, etc. Perhaps most significantly, an incredible and bewildering variety of things were available online for free.

During that time, some companies experimented with unusual and catchy stunts, such as The Spot, a Web soap opera created by a unit of Grey Advertising (now on its own as American Cybercasting); Interactive Imaginations’ Riddler.com Web brain-teaser and scavenger hunt; and the low-tech but homey Mama’s Cucina at the Ragu site. Other companies developed cybermalls with virtual goods on virtual shelves to be placed in electronic shopping carts, but by and large these are antiseptic, boring places to go. Still other Websites built conversation spaces using Web-based chat-room and bulletin-board technology, but few companies knew what to do with the conversations or the participants. A few sites, notably Time/Warner’s Pathfinder, invested heavily in technology and advertising. They reaped high volumes of visitors, but it’s not clear whether their business models will be successful.

A comic note: Everything we’ve just described happened in the past year and a half.

Now companies are spending considerable time and effort trying to deliver to users more of what users want, when they want it. Personalization is hot. Increasingly, users can set keyword triggers and stored queries on various servers. Search engines and directories are moving into each other’s markets, creating vertical services and adding features. Companies such as BroadVision and Ikonic offer sellers sophisticated personalization features with their One-to-One and Customer Intimacy systems, respectively.

Term limits for agents

A discussion of collaborative filtering systems would be incomplete without addressing software agents, an oft-maligned and misunderstood term that spans a wide range of features and functions. At one end, too many things described as agents are nothing more than glorified scripts, macros and stored queries. At the other end, early AI visions and more recent marketing pitches, such as John Sculley’s Knowledge Navigator video and General Magic’s early descriptions of its agents’ incredible
feats, lofted the word "agent" from relative obscurity into high orbit. The promises never materialized, and the concept still hasn’t recovered.

The info-glut has revived interest in agents and software assistants of all sorts. What better way to manage all the bits than with hired help? Vendors of the simpler helpers, such as Farcasts’s droids and NETbot’s bots, don’t use the "A" word at all. Several are positioned as narrowly focused assistants, such as Wildfire’s phone helper and Microsoft Office Assistant, an animated paper clip due to arrive with Office 97. The more complex ones use several interrelated dimensions, including:

- **IQ.** Automation doesn’t take much "intelligence"; making decisions in complex situations without complete information does. Sophisticated agents can perform basic programming constructs, such as looping and conditional branching.

- **Ability to learn and adapt.** Even when agents are carefully programmed to infer preferences from user behaviors, it is often difficult for them to do the right thing. The feedback loop between agent and owner is paramount.

- **Degree of autonomy.** Few agents perform well when the direct feedback loop is missing, which happens when agents run off to do things. IQ, adaptability and autonomy are hard to do all at once, which got General Magic’s Telescript agents in trouble. Its asynchronous agents perform their activities as smart e-mail, which means that they had to be programmed for all contingencies ahead of time. Given how little structure there is in life, that’s a tall order.

- **Ability to communicate.** When agents meet to do business or schedule events for their owners, how do they communicate what they need and have? This is a particularly thorny issue for agents from different vendors.

- **Perceived and actual locations.** Many agents run on client systems, at their owner’s virtual elbow, watching and helping all the time. Others run off occasionally on errands, such as assembling a personalized newspaper by scouring Websites and news feeds. Where the user perceives the agent to be located and where the software actually executes can easily be different. The greater an agent’s IQ, the more likely it is to be a shared resource on a centralized system.

People confuse agents and collaborative filters all the time, partly because of the way researchers and developers use the terms. Agents sometimes use collaborative filtering technology (e.g., NetAngels). Conversely, developers can position filters as user agents (note Firefly Network’s original name: Agents, Inc.). Agents generally do their work for and "near" individuals; collaborative filters run on servers and work for groups large enough to collect a critical mass of reviews.
judgments and knowledge of their local market, and vice versa -- worldwide. Because ACF systems don't merely average all the ratings together, these populations and taste groups should remain distinct and offer increased value to all.

Trial is relatively easy to generate. Curiosity leads many people to rate some items and see whether the system can recommend something worthwhile. To shorten the time it takes to model a particular user's preferences, most systems present carefully chosen sets of items to rate. The first 10 items might cover a broad range of genres; the next 10 or 20 might help distinguish which of several affinity clusters the user probably belongs to. None of the systems yet asks what radio stations or DJs you like, or other questions that might offer quick guidance.

Generating activity

For storefronts, real and virtual, collaborative filters offer a way to generate friction that turns into transactions. Credible recommendations are better than bestseller lists, particularly when they suit a particular person. They also get people talking, even if only to try to figure out why on earth the system recommended a particular item.

Although collaborative filters do perform better than random recommendations, it is unclear how much more accurate they are than more traditional statistical methods such as simple correlation analysis. There is no common measure of performance for ACF systems, and the companies selling them tend to pitch features, not performance specs. Of course, it's hard to compare numbers when there aren't any large-scale systems to measure. ACF systems do many things better than simpler methods, such as handling large information spaces and extrapolating recommendations where few opinions exist.

Even the most pessimistic view that collaborative filters are just efficient high-tech ice breakers gives them an important role. Technophiles may pay more attention and take action when sophisticated technology says they may like something. If it works over time, all the better. Humanists may have an easier time introducing themselves to others if a system that purports to model other people's choices says they probably have something in common. Perception of a match is as important as a good match.

Perhaps most interesting are the business issues surrounding this emerging industry. Scale, flexibility, integration and licensing terms are likely to have as much influence on buying decisions as recommendation accuracy. More on that in a moment.
THE FILTERS

We're happy to see collaborative filters get so much attention: It means academic research is successfully making it to market. Most of the filters described here started as graduate-student thesis projects, often to ease the burden of tracking the popular Usenet newsgroups. Now, of course, the startups protect their algorithms as trade secrets.

Filter tips

Here's how a simple, generic collaborative filter works. The heart of the system is a recommendation engine that can store and manipulate large numbers of vectors that represent reviewers, items and ratings, much as text tools categorize and measure similarities among texts. Most of these engines use several correlation and pattern-matching techniques (a brief taxonomy/genealogy of collaborative filters follows the company descriptions in this section, as a summary, on page 16).

Preserving a participant's identity from one session to the next is critical or the correlations don't persist, so all collaborative filters have some sort of registration process. The next step is to get data into the system, which requires a combination of elegant data-entry design and judicious selection of items to rate in order to maximize their effectiveness.

Participants are probably willing to spend more time with one of these services than they would spend learning a new video game, but not much more. Getting to a recommendation that offers hope of better things to come, quickly, is essential. If the system never seems to converge on a user's likes or dislikes, the user has no reason to keep trying.

Feedback mechanisms range widely, from thumbs-up and thumbs-down symbols to rating bars, drop-down pick-lists and simple lists of items. Because participants often have to rate many items up front, speedy server response is also crucial.

What wine would go with salmon?

A system needs to have a critical mass of reviewed items in order to offer useful recommendations. An ideal database of reviews has plenty of overlap between participants and a shared core that practically everyone has rated -- very likely the opening sets of ratings, which ought to span the breadth of the domain.

It takes several steps for the system to make a recommendation. First it looks for people whose tastes are similar to yours, which means they have rated many of the same items and given them similar ratings. The system then looks for and recommends items that your "trusted neighbors" rated highly that you haven't rated yet.

The hardest technical challenge is maintaining huge datasets and doing the math quickly and inexpensively enough to be useful.

Basis for comparison

Collaborative filters are after an elusive goal. Human recommendations are fallible and inconsistent. If asked the same set of questions several...
times, people are likely to respond with somewhat different preferences or sets. Also, people often use the same words in different ways. That's a problem that ACF systems can tackle directly. They avoid semantic confusion by looking for patterns in the way people rate the same items, independent of how the items might be labeled.

The fuzzy nature of collaborative filter recommendations is accompanied by a lack of measures of performance. Text retrieval vendors, for example, measure the precision and recall of their systems on known document databases. For collaborative filters, Surf Communications' cto David Anderson has created a method that involves measuring the residual error of a series of inverse rating estimates. Put another way, assume you know how a person has rated 100 items. Then remove one item at a time and use the remaining ratings to predict how that person would rate the missing item. Do all of them and sum the error terms to get the measure of accuracy.

We leave this exercise to the reader so that we may present the most notable collaborative filter-based companies.

---------------

Hal Varian, dean of the School of Information Management and Systems at UC Berkeley, has created a superb collaborative filters resource site at http://www.sims.berkeley.edu/resources/collab.

---------------

FIREFLY NETWORK: FINDING NEW FIELDS

Firefly Online, the Website developed by Firefly Networks (formerly Agents, Inc.) is an organic, social Web experience that happens to include sophisticated collaborative filtering technology. The Website, which is a source of music and movie recommendations, is not intended to be a polished, finished service. Firefly positions it as a showcase for the technology the company is licensing to others.

Despite a pretty primitive interface, Firefly Online gets many things right, particularly the melange of social features and functions. For example, Firefly always gives you a sense that other members are nearby -- a relatively simple feature that most Websites lack. When you log in (after registering), the first line after your personal greeting tells you how many other people are currently accessing Firefly. As you move from page to page, Firefly notes which other users recently visited the same pages.

There are many nifty design details to highlight the social nature of the site. As one would expect, Firefly encourages members to create their own template-based Web pages to describe themselves and list their favorite music, movies and Websites. But the list that goes first is of their favorite Firefly members. When you visit another member’s page, Firefly inserts a section in the page called, "Some things we have in common," where it lists any items you and the other member have both listed as favorites.

Release 1.0 19 November 1996
Firefly has a bit of everything, including busy chat areas (the "buzz" section), member music reviews, local mail, hosted zines (called venues), a newsy section (Flypaper) and the aforementioned member Web pages.

Despite the nifty features, Firefly Online’s actual interface design is rudimentary, mostly due to benign neglect. It’s often hard to figure out what is where. The site uses an awkward, drop-down pick-list structure for rating items and telling Firefly what to do. The Firefly team has neglected its interface because it is focused on developing new markets and its server suite. More on that in a moment. Firefly’s lack of slickness and organization has probably been a feature for its users.

There are some important things missing on Firefly, such as direct links to playable music samples and links to a deep inventory of albums for purchase, features that its competitor Tunes.com has (see page 13). Other music sites such as N2K’s Music Boulevard or CD-Now are simply stores.

When I was a wee grub...

Firefly started as the work of several students in Pattie Maes’ Software Agents Group at the MIT Media Lab (for details, see page 16). In March 1995, Maes’ students joined forces with Nick Grouf (now the president and ceo) and David Waxman to found Agents, Inc. They launched the Firefly Website in January 1996, focused on offering music recommendations. It has since broadened to include movies, and now counts an estimated 500,000 registered visitors.

Aware that it must think big and move quickly to stake its claim in the cyberspace attention-making market, Firefly is on a buying spree. It recently bought a news-tailoring product called CRAYON (CReAte Your Own Newspaper) and hired the Media Lab student who wrote FishWrap, another custom-news project. Firefly is considering the purchase of NetAngels, which would complement its capabilities well.

Thinking bigger

The company is loaded for bear. Last month, Firefly Networks raised $10 million in a private placement from Goldman, Sachs, bringing its total outside funding to $18 million. Previous investors include Atlas Venture, Dun & Bradstreet Enterprises, Merrill Lynch, Softbank and Trident Capital. Firefly’s key executives have experience at Alex. Brown, AOL, General Magic, McKinsey, Ogilvy & Mather, Oracle and Powersoft.

Firefly recently licensed its technology to Yahoo! (to create My Yahoo!), ZD Net, Rolling Stone and Reuters New Media. It is moving to use the filtering technology in other ways (e.g., matching products to people), on other platforms (e.g., handhelds) and in new domains (financial services). One thing that may cause prospective licensees to pause before signing is

---

1 Since then, the department has developed many other agents and filtering tools, including Yenta, a matchmaker; remembrance agents; Kasbah, an agent marketplace; Letitzia, a Web-browsing advisor; and Amalthaea, an information-filtering system.
Firefly's revenue model: It is looking for a share of its licensees' revenues, rather than a fixed licensing fee. Firefly is also pitching its services to owners of private networks, as a way to share and navigate corporate information (the infamous Intranets and extranets).

The company's mission is to be "the leading medium for the trusted exchange of reference data." Such trust is hard-won, which means it is valuable. For Firefly that means at least two things. First, well-trained (and thereby trustworthy) filters have built-in switching costs. They are "sticky,"\(^2\) which is good news for companies that want to reduce subscriber churn or increase repeat Website visits.

Second, it means Firefly must take an aggressive stand on privacy, which it is doing. The company has joined eTrust (see page 19) and it has hired Coopers & Lybrand to do audits of its privacy systems and procedures. Partly this is a way to gain the moral high ground in what looks to be a year or two of controversy and debate about online identity management and privacy rights (which we will cover in the next issue).

Five easy pieces

Finally, Firefly is creating an orderly set of software tools to license. The five key elements, available separately, are:

- **Firefly Passport Office.** A portable registration product that allows members to move across sites that use Firefly technology while preserving their preference information. Sites gain immediate access to demographic and other information, without a separate registration process. Members present their passports to "customs agents" on other servers.

- **Firefly Community Builder.** The variety of communication functions described previously, from personal Web pages to chat and mail.

- **Firefly Catalog Builder.** The collaborative filtering engine, which can be used to connect people with products, products with people and people with one another, all tuned to personal preferences. The engine features Automatic Collaborative Filtering and Feature-Guided ACF (see page 16).

- **Firefly Targeted Content and Advertising.** An API with which advertisers can target users according to their demographics, preferences and more.

- **Firefly Analysis.** Reporting tools to assess what visitors are doing and how well the system is working.

---

\(^2\) Short of the complete failure of collaborative filtering technology, the only scenario that would reduce switching costs is one in which every vendor agrees on a representational scheme for easy interchange of user settings. This seems as unlikely as ubiquitous document-interchange standards seemed ten years ago. (Of course, now the world is moving to HTML and related schemes....)
The company's principal technical challenge is achieving scale. The system today is often irritatingly slow, even with a fraction of the users that Firefly would like to see online. It will be interesting to watch My Yahoo! and other sites that are likely to garner high traffic.

Merger discussions

As we were going to press, NetAngels was investigating the possibility of being acquired by Firefly. The two companies would complement each other well. Firefly has ignored its interface in order to improve its collaborative filtering engine, create business relationships in new domains and create business partnerships with sponsors. NetAngels has focused on the user interface and overall experience, as we describe next. The combined services could completely encompass a user's daily Web activities -- an impressive and somewhat frightening prospect. Caveat surfer!

NETANGELS: OWNING THE CLIENT

Regardless how much effort Website owners put into analyzing visitors' clickstreams, there's one limitation they can't overcome: Activity at their site offers information on only a small portion of a user's overall activities. One system, PC Meter, has already built a business tracking and reporting on cross-section usage information for advertisers.

NetAngels is a server-based software agent that uses some collaborative filtering technology among its many features. All of the information NetAngels can collect allows it to target people in unprecedented ways. In principle, a computer vendor could use the system to look for trendy, cheap, educated risk-takers who have visited a computer Website or clicked on a computer-related banner in the past 10 days.

Follow the clicks

Our favorite feature in NetAngels' system is the thin control bar it inserts across the top of every Web page you view using the system. The control bar has thumbs-up and thumbs-down buttons, with which you can easily offer direct (albeit binary) feedback on any page.

NetAngels is smart about how it uses Windows 95, the only platform it currently runs on. It shows up at boot time, asking the user to log in; in the system tray, as a halo icon of last resort (in case you've turned every other NetAngels feature off, you can still launch it from there); as a floating palette to help you navigate; and as an extra button bar above the

---------

3 The NPD Group's PC Meter has quickly become the dominant source for cross-section data on households' Internet activities. NPD created a Nielsen-style panel by installing the PC Meter tracking and reporting software on 10,000 households' Net-connected PCs. NPD's main business is providing survey data to advertisers and other groups.

4 Disclosure: Esther Dyson is an investor in NetAngels.
Win95 task bar that is set to display even when your task bar is hidden -- and has a nifty space for ads. As long as NetAngels is active and you're logged into the system, it knows everywhere you go with your browser.

The in-your-face interface

Depending on your point of view, NetAngels can seem solicitous or intrusive. The company counts on appealing to the silent majority of users who are not militant about their privacy or who trust NetAngels to protect it. The key to NetAngels’ staying on the solicitous side is to offer users value, which it does in many ways.

For example, its AngelMarks feature can take over management of your bookmarks -- stored neatly on the NetAngels server. The system can suggest sites you might want to visit, given your behavior, feedback (via a simple collaborative filtering system), stated preferences and questionnaire answers. Did we mention the questionnaire? When you install the software and register for service, it presents a Meyers-Briggs-style personality inventory for users to fill out. It's optional, but it doesn't say so.

NetAngels’ collaborative filtering is not as powerful as the others described here. Its roots are in the Similarities Engine developed by David Whiteis, who supported local-area networks at the National Institutes of Health in Washington, DC, before joining NetAngels, and created the Engine in his spare time (http://www.ari.net/se/). The NetAngels service, which has only been operating for a few months, now has 2500 users. It expects to get useful collective information when membership is above 10,000.

The system can also suggest other NetAngels users’ pages you might want to check out, but it doesn’t manage the identity and privacy issues well. It would be useful if the system allowed for two levels of disclosure, so people could use pseudonyms for first meetings all the time, then disclose more information as they got to know each other.

From cyber-club to cyber-agent

Mark Goldstein, who founded NetAngels in November 1995, was originally interested in creating a cyber-club service that would automatically link people with similar interests. Then he realized the same technology could be used to find Websites to visit and products to buy, which led him to broaden the company's goals and design specs. NetAngels has most of its software development done by a Polish firm called The Polished Group.

According to Goldstein, NetAngels' current implementation is a proof-of-concept. Goldstein believes that there will eventually be only one or two massive networks serving everyone with TV, phone and Net access. Establishing an early presence is critical, hence the search for larger partners.

Goldstein previously founded and ran Reality Online, which built the Reuters Money Network service, which is private-branded by Schwab, Fidelity and others and is now a Reuters subsidiary. NetAngels has raised $1 million in private "angel" funding. It has also inked deals with several content partners that will promote the personalization service, including Travelocity, WhoWhere?, Fourll and Delphi Internet Services.
SURF COMMUNICATIONS: SMART TUNES

Technologically speaking, Firefly’s nearest neighbor is a newly launched Website called Tunes.com, created by Surf Communications. However, beyond the core engine, the two services are quite different. For starters, Tunes.com is a commercial service; Firefly Online is a proof-of-concept site. Yet both are complete in different ways. On Tunes.com, the content is primary; on Firefly, the social interaction is. Neither has made the recommendation engine central. It runs in the background, helping out. Both approaches are useful. One can’t help wishing the two companies would combine their efforts.

Tunes.com is more organized and complete than Firefly, but less funky and lively. On Tunes.com you can tell where things are and what to do. You can sample any of 200,000 30-second music samples, read reviews and buy the albums directly. At any point, you can add your ratings to an item by using a small feedback bar tucked nicely above the music listings. You can also send other members your favorite playlists via e-mail.

Because the system is built around the music database of record (which Surf has Webified), you can follow individual musicians to see who else they’ve played with and in some cases who their principal influences are. Following the links can be quite educational.

Music buffs with an ear for business

Tunes.com is the brainchild of Kamran Mohsenin and his longtime business partner Jacob Maizel, who, in 1988, co-founded Mobius Technologies, the Macintosh display, accelerator and LAN-adapter vendor. In August 1994, they co-founded Surf Communications, which is privately held, out of an entrepreneurially informed love of music. If they were so interested in learning about music and building their music collections, they figured, others probably were, too. The Web offered a great platform to build a service many people could find and use.

Since Tunes.com is supposed to be exhaustive, coverage is king, for both database content and samples. Surf has struck deals with most of the major recording companies (Warner Music, Sony Music Entertainment, EMI/Capitol Music, MCA Music Entertainment and Polygram) and several smaller record labels to post samples of their recordings online.

The company has created software to speed up the process of recording and posting the samples. It expects to have 1 million tune samples available by mid-1997. As the database grows, users will increasingly be able to listen to obscure, hard-to-find tunes. Surf expects to generate revenues by selling CDs and licensing its technology. HotWired is one of the company’s first licensees of its recommendation engine.

Mohsenin needed a technical wizard for the music venture, and he came up with an ace: David Anderson, who was previously a professor of computer

---

5 Surf licensed Matrix Software’s All Music Guide, which claims it covers every piece of music ever recorded and includes music reviews and biographies.

Release 1.0

19 November 1996
science at UC Berkeley. In 1992, while working as director of software architecture at Sonic Solutions, Anderson created an e-mail-based movie-recommendation system he called RARE (Rating and Recommendation Engine). Later RARE evolved into a Website. You can try it out at http://joe.surf.com/movies/.

Under the hood

After joining Surf, Anderson evaluated ways to improve his earlier efforts. He looked at neural-net technology, but found it unable to model problem domains with large numbers of variables and instances, such as the participants, items and ratings for a mass-market recommendation system. Then he created the vector model he uses now, which represents each user and item as a set of numbers and adjusts them to match that person's ratings. The dimensions/vectors don't have names.

Anderson has added many useful features to RARE, including group recommendations (what will the three of us like?) and a nifty visualization module that graphs members, tunes and movies in "preference space." Now he is creating mood-based ratings, so users can look for music that is mysterious, bouncy or sad. Anderson is also working on acoustic analysis that will allow for similarity searches. Chat and other social features are on the spec for release early next year. Tunes.com is a great start. We look forward to seeing it evolve.

NET PERCEPTIONS: KNOWING WHAT YOU WANT NEXT

Steven Snyder spent over a decade in a systems and consulting career track, learning programming at the University of Pennsylvania, then consulting with Touche Ross and the Boston Consulting Group before joining Microsoft, where he managed the company's IBM relationship and later its languages business unit. Then he left to get a doctorate in psychology from the University of Minnesota, and things haven't been quite the same since.

In 1996, Snyder got to know Brad Miller, a University of Minnesota grad student, and two professors, John Riedl and Joe Konstan. The three had developed a collaborative filtering system for Usenet newsgroups. Their method allowed Usenet participants to add ratings of news items using their preferred news readers and with minimal additional keystrokes. The server technology they developed made recommendations based on information that participants provided explicitly as well as information it inferred from their reading patterns.

The team then extended the pilot system, called GroupLens, to calculate predictive ratings of consumer preferences through a process they call Adaptive Prediction. Their goal is to make it easy for Website developers to enhance their relationships with site visitors.

The four founded Net Perceptions in the summer of 1996, along with fellow computer science PhD David Gardiner. Snyder is the president and ceo, Miller the vp of product development and Riedl the cto. Snyder connected with longtime friend Ann Winblad of Hummer Winblad, who brought initial financing of $600,000. The team launched its product on November 15 and will license it to companies that want to add collaborative filtering to their sites.

Release 1.0 19 November 1996
NET DEVA: TRUSTED ROLODEX SHARING

All the systems we have examined so far feature people rating inert items such as tunes, movies or books. What if one turned that attention on other people? Of course, most people aren’t inert, so the system would have to provide both safeguards against inappropriate use of information and incentives for participating and offering information. More than that, it would also have to offer various levels of participation and trust, to help people distinguish a likely connection (you should meet him) from a recommendation (you should trust everything he says). That’s the ambitious system that Duncan Work is designing.

Work is president of Network Development Associates (Net Dev), a loosely aligned network of professionals that is creating Net Deva, which is still in early development.

Net Deva is a collaborative filter of a different sort. Instead of a central server running a set of algorithms to find patterns, it is a distributed system for collecting, linking and automating multiple annotated Rolodexes over the Net in a way that helps people build teams, staff projects, make introductions and close deals. Its architecture includes software agents that act as helpers and intermediaries. Participants download client software that asks them to profile themselves and their contacts. The Net Deva server publishes user profiles as Web pages and manages several kinds of agents.

Personal search agents look for likely partners, suppliers and other resources (e.g., do you know anyone in the Department of Justice? who has done business in Peru?). When they find a likely prospect, they make a request of that person’s gatekeeper agent. The gatekeeper agents manage information access by evaluating the relevance of requests for communication and checking the validity and trust levels of the agents making the requests. Meanwhile, network broker agents help introduce people who the system determines are likely to benefit from meeting.

A delicate balance

If Work can build a user community with the right mix of incentives and trust, Net Deva promises to offer an environment in which people can share information normally considered too personal to offer to others, including people’s capabilities, professional relationships, goals and values. Personal information is available to the network broker, but not to individuals, unless it is specifically allowed by the guarding gatekeeper. Personal gatekeeper agents are able to give access at varying levels of security, based on levels of trust.

By making this information more explicit and available without exposing each of the contributors, Net Deva helps link informal networks of people. Think of it as a management tool for the human networks that Valdis Krebs described in his guest issue (see Release 1.0, 2-96). Like Krebs, Work developed his ideas and software while he was consulting to corporations.

Net Deva is every bit as ambitious as it sounds. There are aspects of identity management that Work has yet to solve, as well as projects outside his immediate view that he has not yet drawn on. To deal with these chal-

Release 1.0 19 November 1996
lenges, the initial implementation of Net Deva is planned for relatively safe environments, such as intranets and controlled-access networks where trust levels are already reasonably established. Then the company will launch the product more generally, offering the client free on the Net.

Work is collaborating with Bill Hutchison and Kenneth Stephens, co-founders of BehavHeuristics, which developed neural network-based airline yield-management systems in the late 80's and early 90's (see Release 1.0, 2-89). Hutchison is now ceo of Behavior Systems, which is developing software agents that can be trained to carry out complex verbal behaviors automatically. Stephens now runs Operant WebSites, which develops for other companies highly interactive Websites that incorporate multiple technologies, including behavior analysis, instructional design and intelligent agents.

The filter family tree

To recap: Xerox PARC started this field with David Goldberg's Information Tapestry project, which was an active collaborative filter. That is, it required participants to send pointers and reviews to each other, rather than forwarding and recommending items automatically, which is what automated collaborative filtering (ACF) does.

Inspired by the Tapestry, several grad students started projects of their own. One thread at MIT's Media Lab, as part of professor Pattie Maes' Software Agents Group, led to what is now the core engine for Firefly. It began with Ringo, an e-mail based collaborative filtering system created by Upendra Chardaman as his 1994 MIT masters thesis. Then Max Metral built HOMR (Helpful Online Music Recommendations), which worked on the Web and added automatic processing and filtering, making it the first ACF system.

Finally, Yezdi Lashkari created Webhound (aka Webdoggie), which added technology to analyze the content of the items it covered, namely Web pages. Webhound was the first Feature-Guided ACF (FGACF) system. Firefly incorporates aspects of all these projects (see page 8).

Separately, UC Berkeley professor David Anderson created a movie-rating system called RARE (the Rating And Recommendation Engine), which used technology similar to HOMR and added many features. RARE is now part of Surf Communications' Tunes.com (see page 13).

While working at the National Institutes of Health in Washington, DC, David Whiteis created the Instant Similarities Engine, which allows visitors to indicate their favorite movies and recommends others based on previous visitors' votes. The system is quite simple relative to the others listed here. It is one of the features in NetAngels' offering (see page 11).

WiseWire, PHOAKS and others

Collaborative filtering systems are sprouting like weeds. Here are several more noteworthy systems.

Release 1.0

19 November 1996
Empirical Media: WiseWire

WiseWire started as Ken Lang’s 1989 PhD thesis (surprise!) at Carnegie-Mellon University. The thesis was a Usenet message rating and filtering project he named NewsWeeder. In mid-1995, on advice from his thesis advisor Michael Mauldin (Lycos’ founder) and others to strike while the Internet iron was glowing white, he founded Empirical Media.

Its first product, the Personal Surfing Assistant, lets users front-end the WiseWire engine with personal feeds Lang calls U-zines. It has been in beta since May and is scheduled for public release next January. The company will offer a basic free service and an extra-fee premium service.

Lang calls the WiseWire system a collaborative neural network, because it combines neural network-based adaptive content filtering with collaborative filtering. What that means in plain English is that, unlike most other collaborative filters described here, WiseWire can make decisions based on items’ content, not just people’s ratings of the content. (Firefly’s Feature-Guided ACF technology can offer ratings on unread items, too, but as an add-on to the filter engine, not a separate capability.)

This makes it possible for WiseWire to offer predictions on novel content that nobody has rated yet, such as news feeds. The predictions of unrated items use the knowledge gained from collaborative filtering. The WiseWire engine is designed to give ratings rapidly, which makes it applicable to real-time feeds, as well.

WiseWire plans to stay out of the way of other collaborative filtering systems by focusing on text-heavy domains, such as Web pages, news feeds, newsgroups and other informational sources where the content filtering adds value. It is likely that other vendors will enter this area, too, but WiseWire has the advantages of timing and specialization.

WiseWire’s Personal Surfing Assistant requires no client-side software. Its primary interface is standard Web browsers, but it will support a variety of other tools, including e-mail, Marimba’s Castanet and Pointcast. As with other systems, users rate what they see, in this case with a colored bar down the left side of the screen. If users like what they see, they should click high on the bar, and vice-versa.

Lang received seed funding from the Internet Capital Group and is now pursuing a second round of investments. His charter advertiser is SmithKline Beecham Consumer Healthcare. Empirical Media currently has 40 employees, most of them engineers focused on making the system scale efficiently.

AT&T Research: PHOAKS

The Cleverest Product Name award may well belong to AT&T Research, which developed PHOAKS: People Helping One Another Know Stuff. PHOAKS finds, classifies, abstracts and tallies opinions that people write in Usenet news-

6 Neural network technology is usually not well suited for large problem domains, but Lang’s algorithms consolidate overlapping interests.

Release 1.0
19 November 1996
groups. Its principal measure of item importance is the frequency with which discussants mention it. PHOAKS is a useful way to navigate the flood of Usenet messages to find people's opinions.

Unfortunately, the system is not as practical as the other models described here, either in the information it discovers or in the interface it offers.

Others

There are many more collaborative filtering systems and simpler recommendation engines. Each Communications' EachMovie uses Each to Each (sic) technology to recommend movies. Affinicast (from Affinicast Corp., with roots in SRI's psychographic modeling) offers some editorial perspectives to bring users in and participate in its filtering system. MovieCritic from Songline Studios asks new visitors to rate a dozen movies to become a member. It is optimized for video rentals or interactive TV. HubNet Communications has added collaborative filtering to its Boston Restaurant Guide, which takes Zagat's concept into the electronic age.
Until recently, it was quite expensive, if not impossible, to track people. Grocery-store consultants tried radio trackers mounted on shopping carts and ceiling-mounted cameras; Nielsen has used sophisticated logs and People-Meters in homes; others have tried retinal tracking in focus-group settings to figure out where people look. We're sure others have been even more inventive, but none of these methods has offered accurate, detailed results.

We still can't tell who read an ad in a print publication, watched one on TV or heard one on the air. And we can't connect that viewing (or hearing) to the actual purchase. A print or broadcast ad is a very indirect call to action. The potential buyer still has to pick up the phone or visit a store.

Online, of course, there is much more detail available, and it's all being processed by machines with perfect recall. Companies can certainly tell when their ads are clicked on -- even if it doesn't happen as often as they would like -- and in some cases that they were read. (CyberGold, which wants to pay people for paying attention to ads, quizzes users about the ads to prove they did the reading and didn't merely click through blindly.) In fact, if the ad is for an information or entertainment product or service that can be fulfilled online, the ad is really a point-of-sale device. It's quite direct.

Companies may claim to be uninterested in people's identities, but it's all too easy to track someone's true name online, unless that person is smart enough to post to the Cypherpunk's mailing list. Avoiding detection is difficult. Most services ask for your e-mail address, with which it is trivially easy to find far too much information. Your identity -- at least the pseudonym you used to create your account -- is available for all to see.

Two edges, both sharp

The online world is clearly a big step forward in the ability to infringe on people's privacy, and in the ability to serve people better. There is a fine line between the two. We are all too likely to strike an unwitting Faustian bargain in which we lose a lot of privacy in exchange for some consideration we never value, redeem or enjoy.

If the info-flood is a virtual deluge, many of the companies described in this issue claim to be building an ark. The developers of these collaborative filtering systems began their research with many different motivations, ranging from coping with too many Usenet messages to helping people navigate cyberspace, using communication systems to enhance the way people share opinions, inventing new market-making environments and building audiences for sale to advertisers.

Regardless why they were invented, these technologies are now being commercialized primarily for the latter reason: to sell audiences to advertisers. For better and worse, advertising is driving this market today.

What needs protection?

It doesn't help matters that our understanding of online identity is still primitive. There are degrees of explicitness, awareness and presence online
that we are just beginning to explore. For example, five years from now we will be using new ways to contact each other electronically that will replace the busy signal and voicemail jail, which are today's default communication interfaces.

What sorts of guarantees and privileges do we want associated with those new processes? Which of those behaviors will be manifest naturally as social norms that need not be legislated; which will not? What will be the tacit agreements and considerations fifteen years out? We need protections, some sort of citizens' Online Bill of Rights, but it's not clear what has to be protected yet.

One effort to address some of these issues is eTrust, an initiative sparked by Frank Fukuyama's lunchtime talk at the PC Forum last March. After the address, EFF's executive director Lori Fena started talking with Progress Software's ceo Charles Jennings and others; soon they had sown the seeds of a new organization and found more participants.

Fena offered EFF as a host organization for eTrust, which is working to create branded symbols of trust in order to foster the development of electronic commerce. One of the group's first initiatives is to establish some guidelines for privacy, which include labeling for informed consent, the key role of security and the role of context in assessing privacy (read about eTrust in more detail at http://www.etrust.org).

Feedback: closing the circuit

The idea that our default setting should be protection from companies prying into our business is both comforting and puzzling. More than anything, it signals that a feedback loop in our market is badly broken. The break occurred after World War II, with the real start of the mass market and our acceptance of the consumer society. Mass media separates. It keeps readers from writers, buyers from sellers and so on.

The fact that ads can act as point-of-sale devices closes one link of the loop by creating a direct connection between buyers and sellers. The fact that buyers can confer with each other and complain in public spaces online closes another link, as we described last month in the context of customer service. When energy flows through the circuit properly, it encourages appropriate civic behavior.

A lot of energy is being spent to keep the loop open -- the feedback circuit broken -- to keep buyers hidden from sellers. Under the guise of protecting individuals' privacy, of "shielding" customers from vendors' aggressions, advertiser intermediaries insist on offering only aggregate information to their advertisers. Generally, this is a good thing, especially in today's commercial climate. But isn't the best prospective customer one who identi-

7 For example, CyberGold tries to gain more benefits for ordinary people by paying them for reading ads, but instead it mucks up the model by creating a buffer zone that separates parties that ought to be talking. It also ends up selecting members by greed and time available, not intent to buy or demographics.
fies herself? Shouldn't people be willing to be identified in exchange for being treated well? Does the buffering keep us from establishing more direct relationships?

Good treatment doesn't have to mean sweepstakes. Sometimes it means keeping out of someone's mailbox until the time is right, or offering useful software that sports the company's logo across the top. And sometimes it means being available to give good advice when it's needed. This is an opportunity to re-establish relationships that have been broken for years.

Direct contact is not inherently evil. It's only unpleasant when the parties are irresponsible. If companies adopt eTrust labels but they all offer services under the label that represents the same old intrusive approach, the initiative may fail us, even while it succeeds. The key is to make sure that eTrust creates more choices for people, rather than perpetuating the status quo.

The systems described in this issue are characterized by ingenious use of feedback, a feature that we hope appears in many more places. But that feedback isn't being used well. The fact that the feedback in collaborative filtering systems goes only to their central servers and that the service provider promises not to divulge individuals' identities keeps companies artificially separated from their audiences. If we can use these systems to rate what we like and what we hate, why not apply them to the very products and services we use every day? Collaborative filters could be the ultimate focus group.

-----------------------------

**COMING SOON**

- Labels and disclosure.
- Collaboration and conversation tools.
- Navigation.
- The analog world.
- And much more... (If you know of any good examples of the categories listed above, please let us know.)

---

*Release 1.0* 19 November 1996
RESOURCES & PHONE NUMBERS

Lori Fena, **Electronic Frontier Foundation (EFF)**, (415) 436-9333; fax, (415) 436-9993; lori@eff.org

Ken Lang, **Empirical Media**, (412) 422-8450; akl@empirical.com

Pattie Maes, **Firefly Networks**, (617) 253-7442; pattie@agents-inc.com

Max Metral, **Firefly Networks**, (617) 234-5407; fax, (617) 234-5414; max@agents-inc.com; www.firefly.com

Duncan Work, **Net Dev**, (301) 620-8616; duncan@netdev.com

Steven Snyder, **Net Perceptions**, (612) 931-7091; fax, (612) 931-7087; steven@netperceptions.com

Mark Goldstein, **NetAngels Inc.**, (415) 442-0990; fax, (415) 495-5407; markg@netangels.com

David Anderson, **Surf Communications**, (510) 845-9854; anderson@surf.com

Kamran Mohsenin, **Surf Communications**, (510) 649-4704; fax, (510) 649-4706; kamran@surf.com; www.tunes.com

Hal Varian, **UC Berkeley**, (510) 642-9980; fax, (510) 642-5814; hal@sims.berkeley.edu

Hal Varian’s resource page on collaborative filters is at http://www.sims.berkeley.edu/resources/collab.

Except as noted otherwise, all companies’ Websites are at the likely address, http://www.domain_name.com.

---------------------------

**Release 1.0** is published monthly, except for a combined July/August issue, by EDventure Holdings, 104 Fifth Ave., New York, NY 10011-6901; (212) 924-8800; fax, (212) 924-0240. It covers PCs, software, the Internet, computer-telephone integration, online services, groupware, text management, connectivity, messaging, wireless communications, intellectual property law and other unpredictable topics. Editor: Esther Dyson (edyson@edventure.com); publisher: Daphne Kis (daphne@edventure.com); managing editor: Jerry Michalski (spiff@edventure.com); circulation & fulfillment manager: Robyn Sturm (robyn@edventure.com); administrative assistant: Susanna Stromberg (susanna@edventure.com); executive assistant: Helen Martin (helen@edventure.com); editorial & marketing communications consultant: William M. Kutik (kutik@edventure.com). Copyright 1996, EDventure Holdings Inc. All rights reserved. No material in this publication may be reproduced without written permission; however, we gladly arrange for reprints or bulk purchases. Subscriptions cost $595 per year, $650 overseas.

Release 1.0 19 November 1996
RELEASE 1.0 CALENDAR

1996

December 9-10  The Knowledge Advantage - San Diego. Sponsored by Ernst & Young Center for Business Innovation and the Strategic Leadership Forum. With John Seely Brown, Peter Drucker and Stan Davis. Call (617) 722-0320; peggy.quinn@ey.com.

December 9-13  @Internet World '96 - New York City. Produced by Mecklermedia. Keynotes by Lou Gerstner, John Gage and Marc Andreessen. Includes the Internet Advertising Bureau Fall general meeting. Call 1 (800) MECKLER; fax (203) 226-6976; fiw96@mecklermedia.com.

1997


February 8-12  MILIA '97 - Cannes, FRANCE. Sponsored by Reed Midem Organisation. Contact Diana Butler or Pamela Dolan by fax, (212) 689-4438.

February 9-12  DEMO '97 - Indian Wells, CA. Sponsored by Apple Computer, AT&T, Gemini Management Consulting and HP. Now it's Chris Shipley's picks. Call (800) 633-4312; fax (415) 286-2750; mark_davis@infoworld.com.


March 11-14  *CFP '97: Commerce & Communities - Burlingame, CA. Hosted by ACM, Stanford, and UC Berkeley. Seventh conference on computers, freedom and privacy. Send inquiries and suggestions for conference content to cfp97@cfp.org.

March 23-26  **PC (Platforms for Communication) Forum - Tucson, AZ. The twentieth annual; sponsored by us. The living web: models and metaphors. You read the newsletter; now meet the players. Invitations enclosed with this issue. Call Daphne Kis, (212) 924-8800; fax (212) 924-0240; daphne@edventure.com; www.edventure.com.

* Events Esther plans to attend.
@ Events Jerry plans to attend.

Lack of a symbol is no indication of lack of merit.
The full, current calendar is available on our Website (www.edventure.com).
Please let us know about other events we should include. -- Susanna Stromberg

Release 1.0 19 November 1996
SUBSCRIPTION FORM

Please enter my subscription to Release 1.0 at the rate of $595 per year in the U.S. and Canada. Overseas subscriptions are $650, airmail postage included. Payment must be enclosed. Satisfaction guaranteed or your money back.

Name__________________________________________

Title__________________________________________

Company________________________________________

Address________________________________________

City________________________ State__________ Zip________

Country________________________________________

Telephone________________________ Fax____________

E-mail________________________ URL____________

☐ Check enclosed.

☐ Charge my

☐ American Express ☐ Mastercard ☐ Visa

Card Number________________________ Expiration Date__________

Name on Card________________________ Signature____________________

☐ Please send me information on your multiple-copy rate.

Please fill in the information above and send to:

EDVENTURE HOLDINGS INC.
104 FIFTH AVENUE, 20TH FLOOR
NEW YORK, NY 10011

If you have any questions, please contact us at 1 (212) 924-8800; fax 1 (212) 924-0240; e-mail us@edventure.com; www.edventure.com

Daphne Kis
Publisher