

Features Editor: Rebecca Deuel

Multimedia Search: Ready or Not?

Greg Goth

In November 2003, America Online bought the Seattle-based multimedia search engine Singingfish (www.singingfish.com) from global media and entertainment conglomerate Thomson for an undisclosed sum. However, whether the purchase marked the immediate beginning of the mainstream era for non-text-based search or was instead a long-term gamble in a market still getting used to the intricacies of text-based engines such as Google, remains an open question.

"The industry isn't very far along at all," says Jay Webster, president of Fathom Online, a San Francisco-based full-service advertising agency that focuses on keyword-driven advertising media. "If you just look at image search technology, in terms of it being a product, it relates to the question of 'What's the killer app?' How do you apply this stuff to human behavior?"

"I know there have been some attempts to develop things like shopping engines based on imagery, like dragging a picture of a favorite shirt into a box on a search site, and it brings back similar images. But do people really shop that way? I think people will continue to type in the box that they want a striped shirt. It's going to take a tremendous amount of effort to change people's behavior. We're right at the beginning."

Arguably, it won't be just end users who'll need to extensively modify their behavior to bring the age of multimedia search to full flower. From the researchers developing new speech, video, and image search software to workers encoding multimedia files for search

engine indexes, industry observers say the learning curve may be steep.

DEVIL'S IN THE METADATA

One of the most difficult examples is indexing files of news clips, speeches, and lectures and making them more easily searchable.

"On one hand, this problem is much more difficult than traditional speech recognition technology for dictations," says Alex Acero, manager of Microsoft's speech research group. "In a dictation environment, you have several things that play in your favor. One is, you can enroll the system to learn your voice. That may not be possible when you're trying to index speech from who knows who. So your accuracy can go down because of that. Second, when you use a dictation system, you can talk close to the microphone, so your audio signal is nice and clear. But when you record audio data in other environments, it may not be so clear. It could come from a telephone, a battlefield, a room full of announcers, so that's more difficult. Transcribing audio data in those environments means the probability of more errors is greater."

Even for data types (such as audio, video, or still images) that don't depend on speech recognition for a successful search, the most advanced multimedia engines can still have trouble delivering content that an end user wants and that a content provider will consider economically viable and make available. For example, the Web developer between the content's originator and the user might not know how to make the data easy to find or assure that the data's correct.

"Most of the multimedia content out there right now is not really very searchable," says Dan Hendrick, director of search at Singingfish. "When I say searchable, the way we retrieve things from our search index is still by keywords, by text. Most multimedia content out there does not have a lot of useful text around it like you find with a Web page."

Moreover, Singingfish CEO Karen Howe says many Web developers fail to include an adequate quality and quantity of metadata describing content to fully exploit multimedia engines.

"Typically, the people in charge of the coding tend not to put in certain fields that would be very relevant to us," Howe says. "So there are probably 70 different fields that could have

good data in them and you find many of those fields being empty or inaccurate or fairly useless information, like the same information repeated time after time for different streams."

NEW OPPORTUNITIES NEED NEW KNOWLEDGE

Howe says she thinks the problem isn't that these Web developers have trouble transferring text-based search coding skills to multimedia search, but that few developers have much experience with search engine technology at all.

"It's all new for them," she says. "Fortunately, there's been so much attention paid to search in the press that it's piqued people's interest. There are orders of magnitude difference in the amount and breadth of people who attend search strategy conferences now from the number you used to see even recently. And new topics like non-HTML search are getting more and more attention.

"As broadband adoption takes hold not just in the workplace but in homes as well, there's now an economic advantage for those who have streaming media to make it more searchable and findable."

At this juncture, however, using these engines for even cursory searches of existing resources reveals multimedia content of uneven quality and quantity. A search on the Hewlett-Packard SpeechBot engine (<http://speechbot.research.compaq.com>) for Buckwheat Zydeco, perhaps the most recognizable contemporary Cajun and zydeco musician, failed to return any hits. In other cases, the quality or correctness of the data retrieved is faulty. Searching for "lacrosse" in a demonstration image search engine powered by LTU Technologies (<http://corbis.ltutech.com>) turned up a series of pictures with a player supposedly tending the goal. However, the stick he's holding is an offensive player's stick, and the goal is tipped at a 90-degree angle from its proper position, so the player is actually guarding its triangular base instead of its six-by-six-foot target area. In this case, a knowledgeable user such as a magazine editor needing an image of a lacrosse game would undoubtedly not use this image.

Beth Logan, a researcher at Hewlett-Packard Labs who has worked on SpeechBot, concedes that the existing technology and encoding can fall short.

"Some of it's useful for some things," Logan says. "One of the problems SpeechBot has, if

a word is not in the dictionary, and probably the word buckwheat isn't, it can't find the audio. There are still problems. It's good for maybe 80 percent of the things you want to do, but the last 20 percent or 10 percent are going to be hard to do. That's always the case with algorithms."

Yet Logan says she thinks the presumed audience for SpeechBot—journalists looking for audio files—will find much of the material the engine provides useful.

"It doesn't have to be perfect."

BIG FISH, LITTLE FISH, AND SINGINGFISH

At this juncture, even industry executives can't predict which companies and technologies will give potential customers the economic advantage, nor if they will remain standalone entities or become part of a bigger company, as Singingfish became when purchased, in turn, by Thomson and AOL.

"I think, and this is only opinion, that as companies like ours come up with technologies that can be deployed across multiple areas within a larger entity like an AOL, that makes great sense," says Howe. "But you also have existing entities out there that already have search components, so they'll just continue to build up what they've already got. Then you have folks like Google that have a practice of developing from within, that may make an acquisition periodically."

Seth Murray, chief executive officer of StreamSage (www.streamsage.com), a search engine that counts National Public Radio, NASA, and Harvard Medical School among its customers, says he can't tell whether his company will end up as a dominator in its market or as a strategic asset coveted by a bigger player.

"It's hard to tell," Murray says. "What we're doing is definitely strategic for a number of organizations. Look at somebody like Yahoo. Audio/visual search is becoming a real focus for them because they realize they have to shift Yahoo from being a dial-up portal to a broadband portal if they're going to be relevant in an age of broadband access, and that really comes down to delivering high-quality audio/visual content to your consumers.

"We've already started to hit the critical tipping point. Twenty-eight million people now

have broadband access. That's as many people as AOL has in total, and that means groups like AOL, Comcast, and Yahoo have to wake up and realize there's a significant revenue opportunity today to serve those customers. I think you're going to see more and more services tailored to delivering content for broadband."

However, the major players in the search market, including Yahoo and Google, aren't revealing their strategies. Yahoo executives did not respond to a request for an interview (yet late in June the SBC/Yahoo DSL home portal for the first time sported an improved interface in which image search is featured prominently as an option). Google is currently in its pre-initial public offering "quiet period" mandated by securities regulators and was unable to comment.

Fathom Online's Webster says the industry's nascent stage almost mandates that bigger companies not spend much time on these projects themselves.

"It's such a niche-y area, and in this economy, trying to make the argument this is something you must pursue—how do you even size the opportunity?" he says.

"Somebody's going to figure out that with companies like LTU and Pixlogic [image search technology companies Webster is familiar with], it'll be easier just to buy them. I think they're going to look and see either that they're generating revenue or that they have a great product and the bigger company sees the revenue opportunity they're missing and will get them cheap."

DIVERGENT MODELS LOOMING

Currently, the business models for multimedia search seem to be following divergent paths. Much of the early spoken-word audio is news clips—led by National Public Radio, which uses StreamSage technology to index data for Google search, and some of its affiliates such as WBUR in Boston, which has indexed numerous shows on SpeechBot. However, Acero says that news sites tend to provide text that will help identify the clip, so this could drive spoken-word audio to a more specialized arena, one where the material isn't transcribed.

One example he mentions is the OpenCourseWare project offered by the Massachusetts Institute of Technology, which offers course material, including lectures, free of charge (<http://ocw.mit.edu/index.html>).

"One possible use for that is if you're trying to learn about a topic and there isn't much out there," Acero says. "There may be some documents, but it might be useful to do a search for that topic and get a lecture from an MIT professor. That would be a good place to start.

"Another case that might be more interesting is indexing the search of meetings. You could argue our conversation, for example, is a meeting. You could record the audio and keep it in a server. At a later point, you could say, 'I was talking to Alex about this two years ago, let me find out what he had to say about this particular point,' and maybe you can't find it in the minutes you have. So you search. That might be useful."

StreamSage CEO Murray says the company is working on delivering personalized newscasts of material culled from various outlets, but he estimates such a product is at least two to three years off. In the interim, the company is banking on its technology's customizability to deliver quality searches in a wide range of uses. The heart of its general-purpose engine, for example, is a contextual-analysis engine that has read through years' worth of the *New York Times*. For its Harvard Medical School deployment, that engine read through the school's textbooks and online medical journals. Once a user starts searching for a specific topic, that engine uses the stored contextual knowledge to narrow the search to the most relevant terms for that particular instance.

Webster says even in the long term, once metadata can be automatically generated from binary files, that specialized applications might still provide a bigger payoff than consumer queries.

"It might be interesting, for example, for a music publisher to take a new artist's demo recording. What if that demo could be fed into a database and through pattern matching to determine if there's another song in the library that meets a certain standard of relevance? You could quickly determine whether this material is original enough to release without risking copyright infringement suits. They could also use it for sampling. You could back that data into a rights management database and find out whether this has been licensed or not to be sampled. These are applications that could be very useful, and they're all high-end applications for vertical industries."

CONCLUSION

But Singingfish's Howe says the increased visibility of multimedia search in general will result in better technology for all uses.

"Everything other people do is going to do nothing but help us in the long run," she says. "In effect, whether other search engines like it or not, everything they do to encourage people to do appropriate metadata for content will help everybody. It floats all boats."

Related Links:

Purchase or Digital Library members log in:

- **Similarity Search in Multimedia Databases** FREE *Proceedings 20th International Conference on Data Engineering* (2004), <http://csdl.computer.org/comp/proceedings/icde/2004/2065/00/20650873.pdf>
- **Image Representations and Feature Selection for Multimedia Database Search**, *Transactions on Knowledge and Data Engineering* (July/Aug. 2003), <http://csdl.computer.org/comp/trans/tk/2003/04/k0911abs.htm>
- **BilVideo: A Video Database Management System** FREE *IEEE Multimedia* (Jan./Feb. 2003), <http://csdl.computer.org/comp/mags/mu/2003/01/extras/u1066.pdf>