
Interview

The growing metadata challenge – An interview with Joseph Santucci by John Horodyski

Joseph Santucci

is the founder and CEO of piXlogic. Before piXlogic Mr Santucci was the founder and CEO of Metaphor Information Systems, a software company that provided AI-based SAAS products in the content management and CAD space to large scale electric utility customers. Earlier in his career Mr Santucci ran advanced technology programmes at the Electric Power Research Institute in Palo Alto. Joe holds graduate and undergraduate degrees in Nuclear Engineering. He also has an MBA from U.C. Berkeley.

ABSTRACT Traditional approaches to metadata creation and use are likely to come under strain as the volume of material and the importance of video in the enterprise grows. Recent advances in the field of automated image analysis and content understanding offer the potential to simplify critical elements of today's Digital Asset Management systems and related workflows. We had the opportunity to speak with Joe Santucci, founder of piXlogic, a leading company in this field, and get his take on the subject. As piXlogic shows, this new software can provide a critically needed level of automation that will help improve the value equation of today's DAM systems and make them more relevant in a rapidly changing world.

Journal of Digital Asset Management (2010) 6, 225–231. doi:10.1057/dam.2010.23

Keywords: metadata; image understanding; visual search; piXlogic

Recent advances in the field of automated image analysis and content understanding offer the potential to simplify critical elements of today's Digital Asset Management systems and related workflows. We had the opportunity to speak with Joe Santucci, founder of piXlogic, a leading company in this field, and get his take on the subject.

piXlogic is an interesting company that has been working for several years on the very difficult problem of creating intelligent software that can see, recognize and understand what is in an image. piXlogic is an In-Q-Tel portfolio company (In-Q-Tel being the venture capital arm of US Government intelligence agencies). As Mr Santucci explains, 'They discovered us in our early days and saw the merits of our approach and its applicability to the complex problem of finding "stuff" when you have large quantities of pictures and videos and no metadata to guide your search. We have had the opportunity to work with several government organizations in the United States and abroad.

Government customers tend to be a very demanding and sophisticated set of users. We have benefited a great deal by working with these customers over the years. It has helped make our software better, stronger and more scalable, which means we are now much better prepared to address the needs of commercial enterprise customers'.

The name of his company says a lot about who they are and what they do. Mr Santucci remarks that if you dissect the name 'piXlogic' you get:

- 'pix' because we are about images, both stills and moving pictures.
- 'logic' because there is a lot of reasoning in our software.
- 'X' is there because on pirate's treasure maps of old, 'X' marks the spot; that is where the treasure is; that is what you are looking for. We are about searching and finding.

There is an interesting intersection between the work they have done and the world of Digital

Correspondence:
Joseph Santucci
piXlogic, 4984 El Camino
Real, Suite 102, Los Altos,
CA 94022, USA

Asset Management. Mr Santucci makes the point that while current DAM systems are well suited for managing metadata about assets, the cost of generating these metadata is high because much of this work involves manual intervention. Unless someone takes the time to create a set of applicable descriptors (keywords, captions and so on) for each image/video file, DAM system users will simply not be able to find and work with those files. While the situation was somewhat manageable in years past, the sheer volume of new files coupled with the increasing importance of video in the enterprise is creating a real challenge going forward. Manual processes simply do not scale under these circumstances. piXlogic has created software (piXserve) that addresses these issues and that can help customers cope with the challenges ahead.

METADATA CHALLENGES

In the DAM community, we are all familiar with how important metadata are to the process. Many enterprises develop and enforce fairly sophisticated systems to make sure that image/video assets are properly tagged and that those tags be accurate, consistent and suitable for the different types of users that will be interacting with the system. It is not at all uncommon to see implementations where an image may include several dozen tags, each describing different characteristics of the asset and its visual content. Templates, taxonomies, ontologies, namespaces, folksonomies...these, among many, are all techniques that have been used to help content creators select or manually input tags, and much has been written about best practices and related technical and organizational issues.

Mr Santucci points out that there are two fundamental assumptions that are embedded in today's processes, and we are beginning to see where these assumptions might be breaking down as the volume of material to be managed and as the number and types of users to be served increases.

The first assumption is 'contractual' in nature. There is an embedded contract between the creator of the metadata tag and the user. The contract is that they will both use the same word(s) to describe the contents of the image. If the creator uses a word that is different from what a user might have in mind at the time the

search is carried out, that image will not be found. We could look at a painting and decide that 'chartreuse' is the closest colour descriptor for that painting, but what are the chances that a user would think of that word instead of the more common 'yellow' or 'green' as search criteria. In this example, the user will not find that painting. Studies show that this creator/user contract is only fulfilled about half the time. This means we are incurring a large cost to come up with metadata tags but we do not have assurance that the effort will pay off. This issue is only going to grow in importance as more enterprises find themselves catering to a diverse population of users (both internal to the organization as well as external users). The likelihood that the implied 'contract' will fail is only going to increase.

Under the very best of circumstances, searching unstructured data is difficult. Searching pictures and videos ratchets this level of difficulty a few notches further. As Mr Santucci explains, this brings us to the second embedded assumption. For text documents, search has moved over the years from simple text-string/keyword matching, to more esoteric text-mining/document-categorization/entity-extraction/faceted search/and, more recently, meaning-understanding-based content discovery and search. The list of software companies involved in providing/developing solutions of this type, either in an enterprise setting or on an internet-wide basis, is very long indeed. Digital Asset Management provides its unique set of challenges in that the amount of text available for analysis is usually much less for an image/video asset than for a text document or web page. This is why, for the most part, the search functionality available in today's DAM systems is mostly limited to simple keyword-based searches and metadata filtering. This creates a challenge because as the volume and diversity of managed assets increase, keywords searching becomes less reliable as a filtering tool. We can get a hint of this problem by looking at the parallel situation on the web. To varying degrees of success, all of the major internet search engines use sophisticated algorithms to try sift through keywords search lists (which return millions of hits, all having the same matching keyword) in an effort to bring to the top of the list results that are more likely to be relevant to

the user. Today's DAM software systems are far from this level of sophistication, so it is an area that will need much more attention going forward. Simple search is easy, but as the volume of assets increases, the situation gets much more complex and we can expect to see failure points in current DAM systems.

In short, Mr Santucci's thinking is that the current path we are on is going to show cracks at larger scale both from an economic perspective (expense of manually input metadata) as well as from a usability perspective (effectiveness of manually input metadata).

A COMPANION APPROACH

What we can do with pictures/videos is to analyze the image itself and derive searchable metadata from this analysis. This gives the hope of, if not replacing, at least supplementing keyword-based searches in DAM systems.

Creating software algorithms that help us automatically search images has been a very intense field of study for the better part of the last 30 years. For the most part, researchers and software vendors in this space have based their development on techniques that try to match patterns of pixels in one image to those in another image. Mr Santucci explains that when these techniques are successful, they generally do so in a narrow sense (a particular type of image, a particular type of scene, a particular type of search and so on). In reality, one can find similar patches of pixels in images that do not have much to do with each other and this can give rise to undesirable 'false positives' when the same techniques are applied to more general situations. Unfortunately, these more general situations are exactly the cases that are of interest to DAM users. For instance, if you were to use signatures such as texture or colour as a figure of merit, you would find little difference between a yellow sun, a yellow t-shirt or a yellow trash can, but of course these are all quite different things. What is needed is something a bit more sophisticated.

In piXlogic, we have a company that has moved significantly beyond this type of analysis and has established a new level for the state-of-the art. Mr Santucci explains that piXlogic understood early on that it needed to address the problem in a more holistic way and tackle the difficult stuff up front. The company's

internally developed core-technology includes three key components: (i) the ability to automatically 'segment' the contents of an image in a way that the segmentation products tend to correspond to what we as human beings would characterize as a 'logical visual object'; (ii) the ability to describe and compare the appearance of visual objects in different images (searching is a process of comparisons after all); and (iii) the ability to reason and understand context about the image and what it contains (what piXlogic refers to as 'notions').

Each of these areas poses a significant set of challenges in its own, and has been the subject of intense work at piXlogic for the past several years. For example, consider an image of a person, where the individual is standing with hands in their pockets. After the logical visual object has been segmented out of the image (for example, the person and its components/subcomponents), the overall shape resembles very roughly that of a pencil, with a semi-roundish extremity on top. This is a simple shape that is easy to compare and detect the presence of in other images. If the same person, however, stretches out their arms and legs, the physical shape will look completely different, more akin to a five pointed star, and not readily comparable to that of a pencil-like object. The logical representation, however, is unchanged as the person is still the same person. piXlogic has advanced the state-of-the art in this field by developing technology that can deal with such seemingly complex types of comparisons.

Being able to segment an image, and being able to make comparisons at a logical-visual-object level are but the starting point to the visual search process. Mr Santucci remarks that much of what we think we see with our eyes in an image is not actually there, and is just the result of reasoning and interpretation by our brains. The piXlogic software also reasons about the objects it sees in a picture, and in some cases it is able to understand what type of object it has just seen. piXlogic calls these understandings 'Notions'. Notions are important because they provide contextual information about the image so that the piXserve software can make some sense of what it sees. For example, the software can understand what is background and what is foreground from a single still image, and this allows users to search

by focusing on one or more foreground objects or by background alone. The piXlogic software includes notions for many everyday life concepts such as: 'car', 'building', 'airplane', 'flower', 'vegetation', 'sky', 'fire', 'sand', 'sea', 'face', 'person', 'bikini' and so on. Without user intervention, the piXlogic software can automatically create an index of the contents of pictures/video frames down to the visual-object level. In addition to this rich set of metadata, when the image/video frame contains something that the software understood then the corresponding label ('notion') is automatically added to the records for that image.

The piXlogic software (piXserve) can be put to use directly as a search solution that is accessed by end-users, or indirectly as a tool to feed metadata to an existing DAM system.

piXserve is a server-based solution and is available in a range of editions (workgroup through enterprise level). More recently, through a partnership with Amazon Web Services, they have also rolled out a cloud version that allows customers to scale software use based on an 'On-Demand' or 'Utility Computing' model. piXserve helps enterprise customers in three critical areas:

- Content Discovery (find pictures/videos that contain specific objects, scenes, text or people of interest);
- Content Auto-tagging (automatically label an image/video);
- Content Alerting (automatically inform users when items of interest appear in a live video stream or a web crawl or just an indexing run).

Creating a searchable index with piXserve is simple, you just point the software to the location on your network (the hard drives, the storage device) that contains your pictures/video files, and the software will just go through, read all those files, do its calculations, and store the results in a database that it will later use to support user queries. The software can be set up in 'automatic' mode, so that it can keep an eye on a repository and update its index as new files are dropped in those locations. There are other ingestion routes that are also available. The software can index live video through multicast-IP connections. Just provide the IP address of the sources to be monitored, choose the

transport protocol, and piXserve will do the rest. Another route that is also available involves crawling web sites. A crawler is available within piXserve, so you can create a list of IP addresses that you want to scan, how many levels deep you want to go, and how often you want to refresh that index.

The beauty of automatic indexing of the type that piXserve provides is that no implied creator-user contracts are required, and the only cost is machine time (for example, it scales). The software is indiscriminate in what it catalogues. If it can see it in the image, it will create a record describing where it is, what it looks like, what is near it, what the context is and so on. How these metadata are used is completely up to the end-user, at the time the search query is formulated (see Figure 1). As Mr Santucci puts it, 'We are not saying that you don't need or won't benefit from manually applied tags, but we are saying that piXserve-style automated indexing will help you fill gaps that manual processes leave open. It shifts responsibility for the success of the query more into the hands of the user, rather than those of the content-creator/archivist/logger, and this is a good thing'.

Users can interact with the software and formulate different types of search queries. Examples might be: (i) find me a picture/video segment where something that looks like this coffee mug is visible anywhere in the image; (ii) find me a picture/video segment where a sign that says 'French Bistro' is visible; (iii) find me a picture/video segment where the software could recognize the presence of the notions 'car', 'building' and 'road' all within the same image/frame (for example, in essence a city-street shot); and so on. That the software can carry out these types of searches using only internally and automatically generated metadata is remarkable, and provides a set of very high value functions to DAM users.

To administrators, piXserve can provide critical assistance when it comes to adding keywords/tags metadata to DAM systems. If the software has indexed a repository of images for which external metadata are available, piXserve can make recommendations about how to tag an image/video that does not have keywords associated with it (see Figure 2). It does so through a process that considers how images that contain similar visual objects, similar

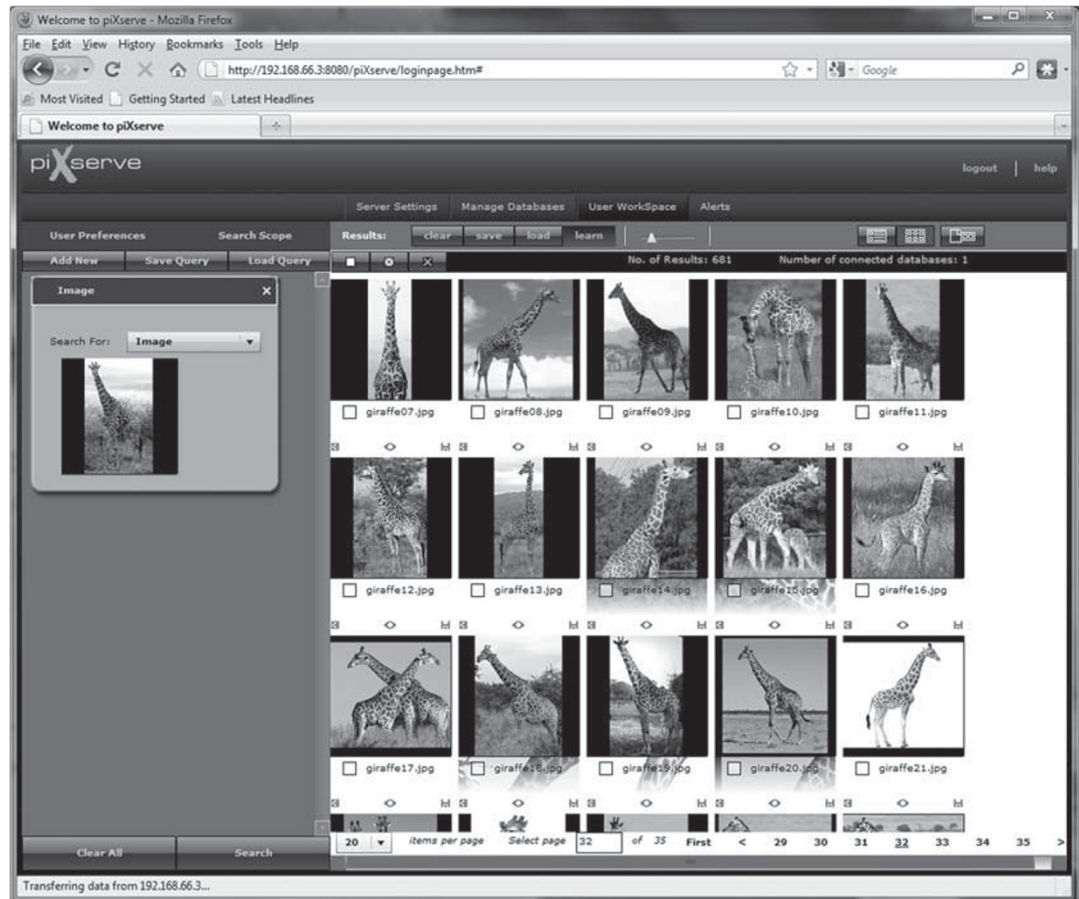


Figure 1: Using an image of a giraffe to search and retrieve a list images that also contain giraffe-looking objects. piXserve can create this search experience through its automated indexing process and without requiring any additional metadata other than the query image itself.

notions and similar external metadata have been tagged in the past. This level of automation can significantly reduce the cost of adding tags in a DAM system. It can also improve quality and check for consistency in how things have been done in the past (ex: I ran the software and now I can see that I have 20 pictures that were catalogued consistently, and here is a twenty-first that has similar content but the tags are different). You can create automated filters that let you flag 'inconsistency situations' automatically. This is exactly the kind of data that lets you discover when/where your internal workflow processes are breaking down so you can do something about it.

In a large system, there will be several piXserve machines that are dedicated to creating and updating indexes of material from many sources. piXserve-ALERT is an add on module

that flips the search process on its head. A user can set an alert criteria and piXserve-ALERT will keep track of what is going on with the piXserve indexing machines. If an image/video comes across that meets the user's alert criteria, a signal is generated. The signal takes two forms. An e-mail is sent to the user, showing the time the alert was generated, the criteria, the feed and a link to the item. You click the link and the image is displayed or the video starts to play. The second signal is JMS based. piXserve is a J2EE application, and as part of the stack there is the possibility to store messages that can be picked up by third-party applications. The message (the alert criteria and triggered result) can be picked up by another application that does something with it (for example, start to record or turn off the lights, or whatever the third-party software was designed to do).

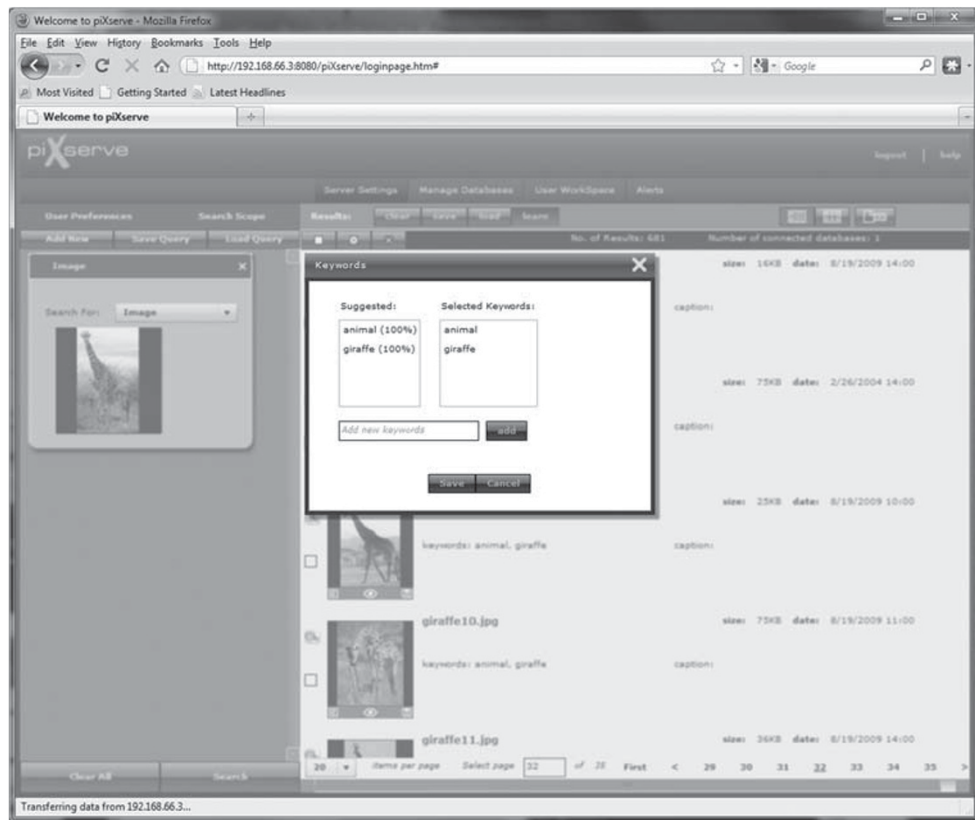


Figure 2: piXserve keyword recommendations. In this example, some of the images of giraffes in the database also have some keywords associated with the image. piXserve can use this information to make recommendations about how to tag an image that does not have keywords.

Alert criteria are formulated the same way a user formulates a search criteria. They can combine one or more images (or different objects from different images), text strings, notions or any of the search modes available in piXserve (GPS location keywords and so on). Up to six can be combined using logical operators (AND, OR, NOT). They are saved by the system, and remain there as long as the user keeps them there. You create one, you submit it, and you walk away. Sometime later (hours, days, weeks, months) when it is triggered, you are notified. If we think of ‘search’ as a ‘lean-forward technology’ (for example, the user is actively engaged on the computer screen, doing searches, saving results and so on), then we can think of ‘alerts’ as a ‘lean-backwards’ technology (for example, I told you what I want, now leave me alone until it comes across the network).

There are many applications for the content-based alerting capabilities provided by piXlogic. In

a commercial environment, it helps create closer connections between end-users and what the organization is doing with respect to producing/creating/ingesting images and video assets.

IMPROVING THE BOTTOM LINE

The ‘notion auto-tagging’ functionality, the ‘keywords recommendation’ functionality and of course the search functionality provided by piXserve, all are critical factors in the workflows of enterprises that manage substantial volumes digital media.

piXlogic is not a Digital Asset Management software provider *per se*, and in fact it was only recently that they started marketing in this space. The piXlogic software can fill a very significant gap in DAM space. As Mr Santucci explains it: ‘DAM systems do a very good job of helping users manage metadata and manage assets based on these metadata. However, the metadata need to be supplied, for the most part, manually. We can help reduce the cost and



improve the quality of these activities because our software can provide desperately needed automation. In addition, when no manual metadata are available, our software can step in and still let users carry out valuable searches. Overall, the software provides significant advantages (speed, cost, simplification) that can augment existing Digital Asset Management systems regardless of how simple or sophisticated they are. Our Web Services API (based on REST) makes it easy to integrate piXserve into existing workflow environments. This is really quite exciting because whether you use piXserve on a standalone basis, or whether you choose to integrate it as part of your overall DAM strategy, we can significantly enhance user experience and workflow. It's about making staff more productive, decreasing costs, and improving revenue'.

Overall, the capabilities that piXlogic brings to the table can affect business strategy in two basic ways.

Internal operations, back-end: Reduce costs and improve staff productivity and quality

- Issue – If you are deploying a digital asset management system you are facing a significant cost for the manual tagging of assets that need to be ingested in the DAM system. This cost can be a big fraction of the total overall system cost. The activity also adds time to your deployment schedule. Further, you also have to worry about quality assurance because different people may choose different tags for essentially similar images. If the volume of material that you are ingesting on a daily basis is large, you are facing a significant and recurring cost for keeping your initial index up to date. If you are dealing with a significant amount of video, the problem escalates much further.
- Impact – The piXserve 'notions based auto-tagging' and 'keywords recommendation' functions will reduce your cost. These functions decrease the amount of work your staff has to do, and at the same time improve the quality of their work product. The savings because of staff productivity can be very substantial. Equally important, you will be saving time. This means that your assets will be available to users more quickly. In today's environment, time is a key competitive factor.

Customer facing, front-end: Increase revenue, differentiate from competitors

- Issue – Your customers search your archives using your predefined set of metadata and this may/may not match how they go about looking for things. When they do not find what they want, they walk away and go to your competitors. Further, you do not have a good way to know what they were trying to find and why they did not find it. This is very valuable information that can be used for the purposes of improving the search experience thereby achieving more product sales.
- Impact – The piXserve search functionality allows users to navigate your web site using example images that they provide (it can even come from other websites). In the case of video content, piXserve lets users more easily get to the right segment of the video that contains what they are looking for. As a result of these features, they are more likely to find what they are looking for, and this drives sales. piXserve can help expose more of your assets to customers, and this improvement goes straight to your bottom line. In addition, piXserve can give you information about what users are searching for (as administrator, you can see their actual examples). This gives you the opportunity to make adjustments and improve sales (refresh content to better address customer needs, or add tags to existing content to make those types of assets more accessible). Users want to work with companies that make it easier to do business with them. Image search is a distinguishing feature that can set you apart from competitors.

SOME CLOSING THOUGHTS

Metadata are the lifeblood of any DAM system. Our traditional approaches to metadata creation and use are likely to come under strain as the volume of material and the importance of video grows. Image analysis-based techniques have matured in recent years, and, as piXlogic shows, they can provide a critically needed level of automation that can help improve the value equation of today's DAMs and make these systems more relevant in a rapidly changing world.