The Open University

More than a thousand words

Stefan Rüger
Professor of Knowledge Media

Inaugural Lecture on 4 June 2008

Abstract. Web search engines have raised our expectation to be able to search for documents in large repositories with just about any query word. This lecture will examine the corresponding challenges and opportunities of Multimedia Search, ie, finding multimedia by fragments, examples and excerpts. Rather than asking for a music piece by artist and title, can we hum its tune to find it? Can doctors submit queries consisting of medical scans in order to identify medically similar images of diagnosed cases? Can your mobile phone take a picture of a statue and tell you about its artist and significance?

Short biography. Rüger read Theoretical Physics at Free University Berlin, gained a PhD in Computer Science at Technical University Berlin in 1996 and carved out his academic career at Imperial College London. In 2006 he joined The Open University's Knowledge Media Institute as Professor of Knowledge Media. He has co-authored over 100 scientific publications during his career. For details see http://kmi.open.ac.uk/mmis.

A video replay of my inaugural lecture in quicktime format is available at http://stadium.open.ac.uk/stadia/preview.php?s=1&whichevent=1167. These are post-inaugural notes — produced from slides and the script for the lecture.

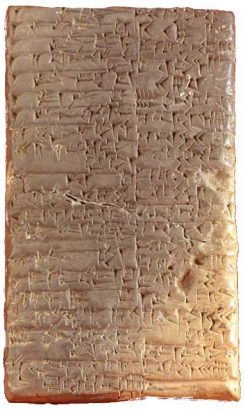
Thank you, Vice Chancellor, for the kind introduction.

Dear colleagues and guests,

Welcome to my inaugural lecture as Professor of Knowledge Media. I always wondered how it would be to give an inaugural address myself, having seen so many colleagues going through this rite of passage.

I found out that preparing an inaugural is easier than some other tasks: just recently I was asked by my 9-year-old son, Ryan, whether I could explain to his class what I was doing at work. He wanted to know whether I could help for the Science, Engineering and Mathematics day of his school. So, I explained to him: "It is quite simple: I try to find out how best to search in a large amount of multimedia — films, images, books and so on." "Do you make the catalogue for Argos?" he asked. "Not quite," I smiled, "but this is actually a good example of the data that we deal with." "What has Science, Engineering *or* Maths got to do with it?" he impatiently asked, continuing, "isn't the Argos catalogue just for finding stuff in the shops? Even grandma finds things in the shops — actually better than you!" So, I figured that an inaugural would be easy to prepare, but dealing with hard-nosed 9-year-olds in school will be much tougher!

A brief historic excursion. Most of my research is about ways of finding media objects in digital multimedia repositories. In some sense this task has existed in one way or another for around 5000 years, and today we still use some of the very same technologies to manage collections.



Slide 1: Clay tablet

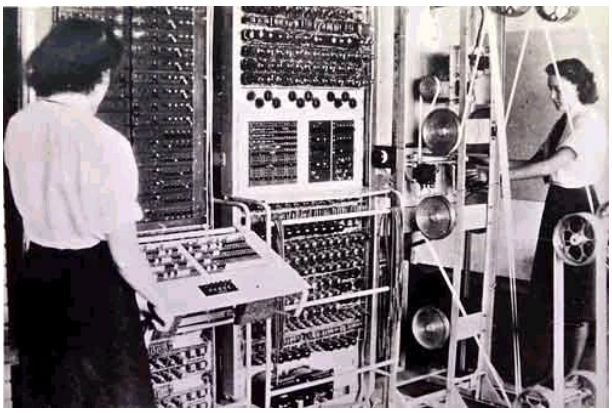
For example, look at this wonderful cuneiform clay tablet from the 24th century BCE from the cradle of civilisation in Mesopotamia.

Hundreds of thousands of these tablets survived war, conflict and burning down of libraries and temples. Actually, the fire may have preserved more of the tablets than was intended at the time! These tablets recorded the business accounts and religious texts. They were usually kept in storerooms full of boxes. The storerooms were often underground and inconvenient to reach, sometimes only by a ladder through a hole in the ceiling. Each of the boxes had a clay tablet that detailed its contents. The Sumerians in ancient Iraq did not have titles on their documents: they used *incipits* instead, which are the first few words of the text. Here is an example of incipits on one tablet (Dalby 1986):

1. Honored and noble warrior
2. Where are the sheep
3. Where are the wild oxen
4. And with you I did not
5. In our city
6. In former days
7. Lord of the observance of heavenly laws
8. Residence of my God
9. Gibil, Gibil [the fire god]
10. On the 30th day, the day when sleeps
11. God An [the sky god], great ruler
12. An righteous woman, who heavenly laws

Those incipit tables served as orientation for browsing and searching 5000 years ago, and still today we use little catalogue cards with a small description of multimedia objects to facilitate finding them.

Of course, the world of information and knowledge has radically changed over time: first through civilisation, then manufacturing, digitisation and the internet: No longer are the only existing written records guarded by priests and scribes in places to which neither you nor I would have had access. During the development of civilisation we have come to recognise that education is one of the most valuable assets for society, and indeed education is free or heavily subsidised in many countries. One prime example is The Open University's OpenLearn project in which high-quality learning material is freely made available on the Internet.



Slide 2: Colossus

Undoubtedly, computers play an important role in this revolution. To set the context, here is a picture of the first semi-programmable electronic computer, Colossus, designed, built and deployed in 1943, literally 4 miles from here [Walton Hall in Milton Keynes] in Bletchley Park: No keyboard — just switches, no monitor — just lights, no memory — just a rotating ticker tape. Its sole purpose was to crack the encrypted messages of the Nazi High Command, the Lorenz cipher, during World War II.

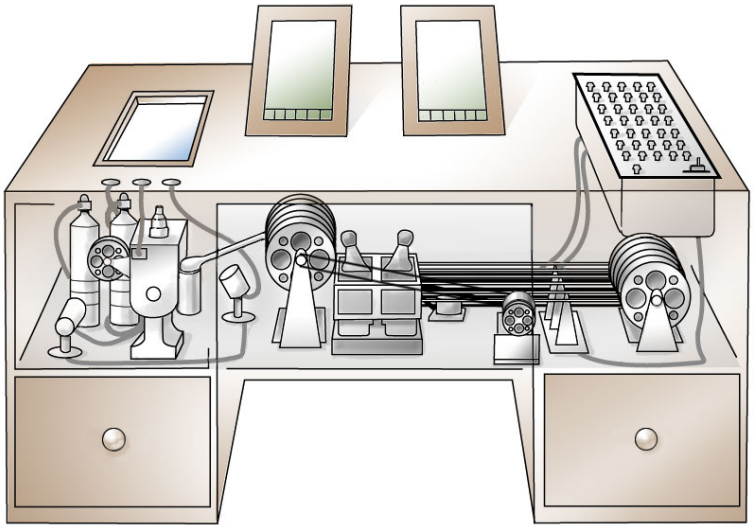
Slide 3: Vannevar Bush 1890–1974

One of the early true visionaries of that time was Vannevar Bush: he predicted the internet and modern digital libraries long ahead of their time — and did not live to experience them! Vannevar Bush was the director of the Office of Scientific Research and Development and lead 6000 scientists in R&D for World War II, including the Manhattan project that built the first atomic bomb. In July 1945 he published an essay "As we may think" in The Atlantic Monthly, in which he described the idea of a memory extender, short *memex*. This fictitious machine (see Slide 4) would be integrated into a desk that contains a glass plate with a camera to take pictures of pages, microfilm storage, projection systems for two screens, and a keyboard with levers. Its purpose was to provide scientists with the capability to exchange information and to have access to all recorded information in a private library that contained all your books, correspondence and own work. It would function as a rapid information retrieval system and extend the power of human memory.

More importantly, his design included the concept of associating resources and adding comments to them. The most remarkable fact of the memex design is that its links are very

similar to the links on web-pages. For that reason Bush is now seen as the grandfather of the world-wide web. It took nearly another couple of decades after the invention of the internet to realise the element of memex that allows adding your own work: wikipedia, an online encyclopedia, was born in 2001, where everyone can link to resources, add comments and corrections and write own articles.

Digitisation means that we are no longer restricted to catalogue cards. We can carry out full-text searches in books and articles with just about any keyword that we deem relevant; computers tell us who the author is, what the title is and where to find the resource. Rather than searching through catalogues to find the books, computers search through books to find their catalogue cards.



Slide 4: Memex design

Increasing computer power and collapsing storage costs make it easy, conceptually at least, to search for text in huge collections. Computers insert *every* of the documents' words into an index that points back to the documents.

The index of any web search engine looks just like the text index from this travel guide (Slide 5) with the differences that every word is indexed and that documents are listed instead of page numbers.

Slide 5: Index of a book

I do not want to hide the fact that there remain a large number of engineering challenges and many other research questions, too: what is the best way to rank documents, how to best utilise document structure, how to best express the user's information need etc. These research questions, and many more, are catered for by the thriving and large field of Information Retrieval.

The most powerful lesson that I learnt over the years is that the best research adds value to whatever was there before. This insight was sparked when I asked the students of my Digital Library class to show hands if they use online bookshops and if they use the university's inter-library loan system.

I should add at this point that inter-library loans are normally free for students and that students have access to virtually all of the 50m books worldwide. This includes the 3m books currently in print, which you can buy via online bookshops. So, to paraphrase, I thought I had asked a no-brainer: Do you prefer access to lots of books for free or would you rather buy from a smaller pool? To my surprise, the overwhelming majority of my students preferred buying books online. When looking closer at the two, the big difference is that online book stores offer a robust search facility, scanned title pages, a table of contents, an abstract, media and customer reviews, personalised interfaces, full-text search, and the perception of fast delivery. None of these added values can the interlibrary loan systems compete with.

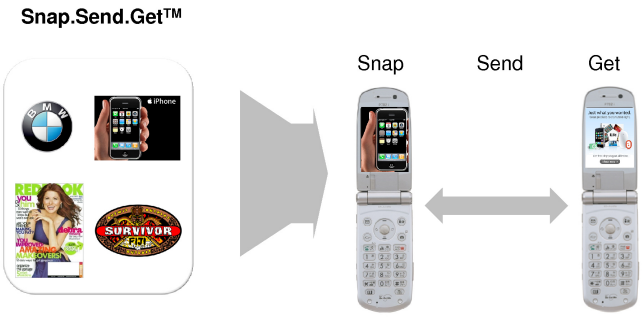
So, the lesson I learnt was that whatever research I undertake in multimedia information retrieval, it should add some value. Today I would like to show a number of tricks of the trade that add value for multimedia retrieval. Most of the research presented here was done at Imperial College London before I joined The Open University.

Slide 6: Milton Keynes's Peace Pagoda

Multimedia retrieval. What is multimedia information retrieval? At its very core it means finding multimedia documents and building multimedia search engines. For these, the query can be text, audio or images: For example, if you walk around in pleasant Milton Keynes you may stumble across this interesting building (Slide 6).

Would it not be nice, if you could just take a picture with your mobile phone and send it to a service that matches your picture to their database and tells you more about the building (in this case *"Built by the monks and nuns of the Nipponzan Myohoji, this was the first Peace Pagoda to be built in the western hemisphere and enshrines sacred relics of Lord Buddha. The Inauguration ceremony, on 21st September 1980, was presided over by the late most Venerable Nichidattsu Fujii, founder ..."*[1]

You see, multimedia search has a wider remit than just text search. Not only can the documents be of any medium, but also the query. Now assuming we could perfectly search with methods like these, what could we do with multimedia search?

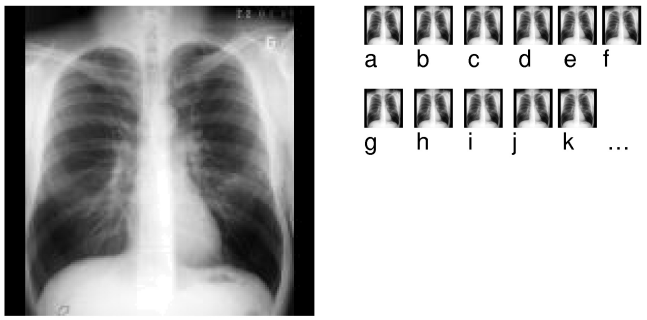


Slide 7: Snaptell's service

There are obvious applications in tourism as seen before. There are also applications for advertising that so much seems to underpin the whole search industry: A colleague of mine, Prof Manmatha of University of Massachusetts at Amherst, is also co-founder and advisor to Snaptell Inc, a startup specialising in mobile image search. Their idea is that customers take pictures from print-media, send them in and receive promotion or product information, vouchers and so on. Just a few weeks ago Snaptell did a major campaign to promote a new movie. Customers sending in a picture of the print poster

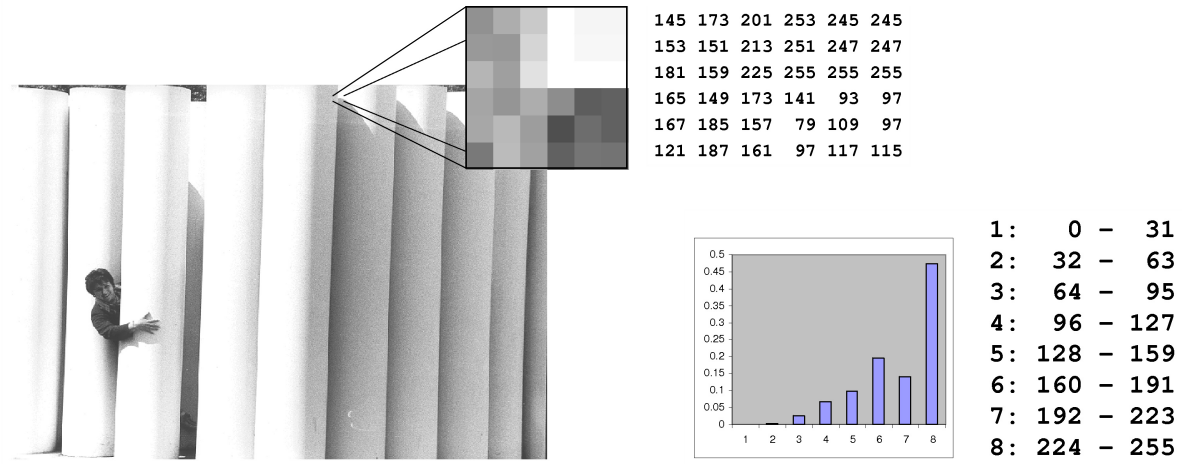
[1] http://www.mkweb.co.uk/places_to_visit/displayarticle.asp?id=411

received an exclusive trailer, saw showtimes and probably could straight away phone to order tickets. One obvious benefit for advertisers is that they receive feedback as to where which print advert was noticed.



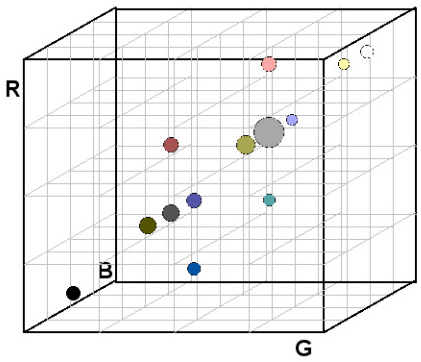
Slide 8: Medical-image retrieval

On a more serious note there are obvious applications for medical image databases. When I come to see a doctor because I suffer from shortness of breath and coughing, the doctor might wonder where she or he has seen the dark shadow on the X-Ray before. If computers are able to match significant medically relevant patterns, they can return data on these cases, so the doctors can undertake a differential diagnosis. Of course, there is still a great deal of difficulty in the multimedia matching, and promising research in some areas has just begun.



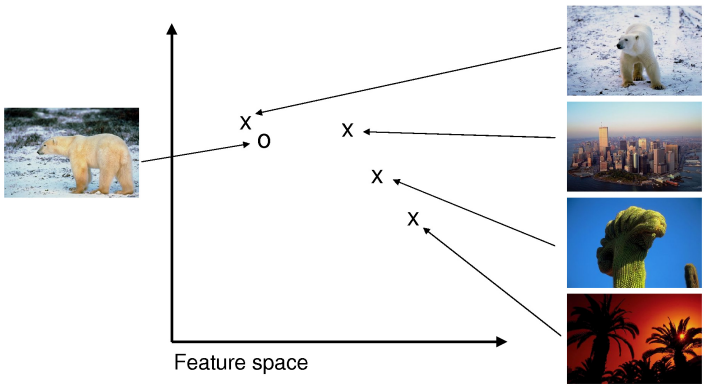
Slide 9: Millions of pixels with intensity values and a simple intensity histogram

Some of the difficulties come about by the sheer amount of data with little apparent structure. Look at this black and white image, for example (Slide 9). It literally consists of millions of pixels, and each of the pixels encodes an intensity (1 number between 0=black and 255=white) or a colour (3 numbers for the red, green and blue colour channel, say). One of the prime tasks in image retrieval is to make sense out of this sea of numbers. The key here is to condense the sheer amount of numbers into meaningful pieces of information, which we call features. One trivial example is to compute an intensity histogram, ie, count which proportion of the pixels falls into which intensity ranges. For this example (Slide 9) I have chosen 8 ranges, and the 8 numbers tell us a rough distribution of brightness in the image.



Slide 10: 3d colour histogram

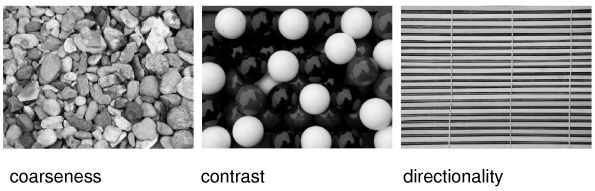
Human perception of colour is three-dimensional and colour histograms are 3-dimensional as depicted here (Slide 10). Again, this diagram shows a crude summary of the colour usage of an image, here the Peace Pagoda (Slide 6). Each of the Red, Green and Blue colour axes in RGB space is subdivided into intervals yielding 4x4x4 = 64 3-d colour bins.



Slide 11: Features and distances

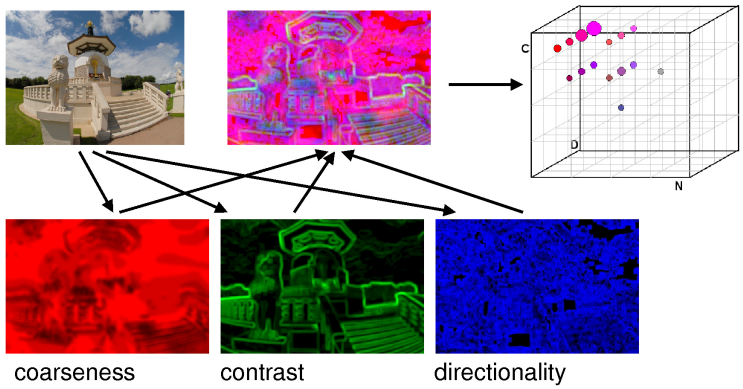
Finding an image by example (when the query consists not of words but of images) requires us to compute the features and use these features to identify the closest match. This slide (11) shows the main principle of *search by example*; in this case the query is the image of an ice-bear on the left. This query image will have a representation as a certain point (o) in feature space. In the same way every single image in the database has its own representation (x) in the same space. The images, whose representations are closest to the representation of the query are ranked top by this process. The two key elements really are features and distances. Our choice of feature space and how to compute distances has a vital impact on how well visual search by example works.

Of course, the features of Slides 9 and 10 are very simple, and the features that we compute are normally more complex than that. For example one of my former PhD students, Peter Howarth, studied and devised ways to extract texture descriptions from images. Psychologist have found out that we humans respond best to coarseness, contrast, and directionality (Slide 12, Howarth and Rüger 2005).



Slide 12: Textures

Unlike colour, which is a property of a pixel, texture is a property of a region of pixels, so we need to look at an area around a pixel before we can assign a texture to that pixel. Slide 13 is an example, how we compute for each point in an image (by considering a window around this point) a coarseness (C) value, a contrast value (N) and a directionality value (D). These values can be assembled into a single false-colour image, where the red, blue and green channels of an ordinary image are replaced by C, N and D, respectively. This expresses visually the use and perception of texture in an image. From the false-colour texture image we can then compute 3d texture

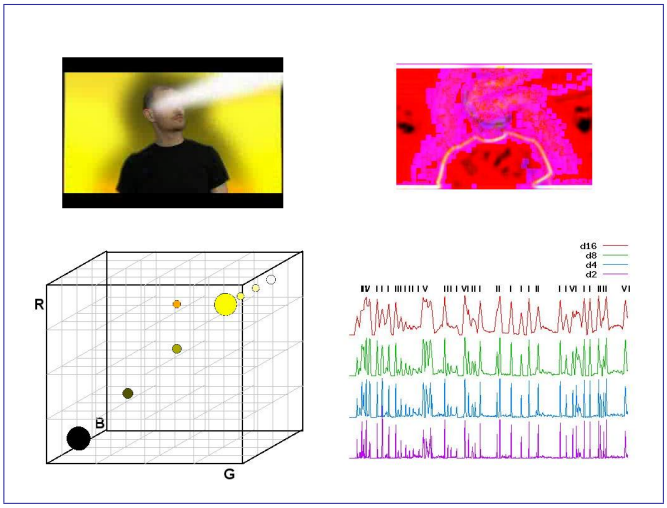


Slide 13: 3d texture diagram via false-colour images

histograms exactly in the same way as we do for colour images.

Slide 14: Anticipation, a SciFi trailer by Vlad Tanasescu (click frame for video)

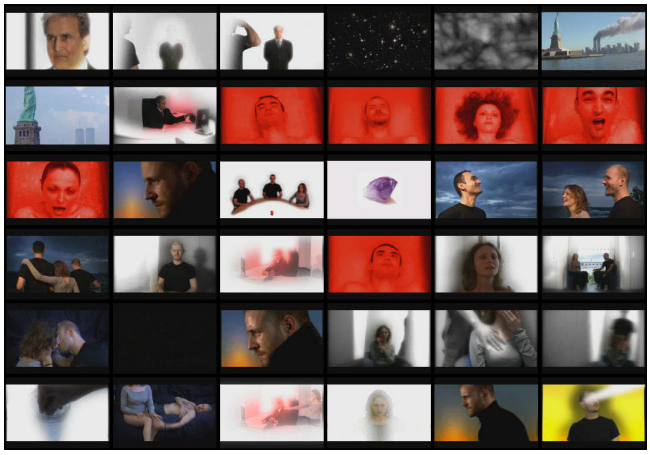
At this point I would like to show how features can help processing videos. One of the first tasks for indexing any video is to cut it up into the shots, ie, the segments that were used to edit the movie. Normally the change from one shot to another would be either an abrupt change (a cut), or a gradual change, for example, by fading in and out. In order to detect shots it is useful to compare neighbouring frames to see whether they are significantly different. For gradual transitions it may be useful to additionally compare frames that are further apart (say, 4, 8 and 16 frames). I will demonstrate this technique with a short movie that was made by Vlad Tanasescu, a PhD student at KMi. The movie is a science fiction trailer. First I invite you to watch the video by clicking on Slide 14 (or via http://mmis.doc.ic.ac.uk/inaugural/Anticipation.mpg).



Slide 15: Processing a video to detect shot boundaries (click frame for video)

This video stream is decomposed into single frames, and each frame is subdivided into a 3x3 grid of tiles. For each tile we compute a straightforward 3d colour histogram as in Slide 10. Then we define the distance between any two frames as the median of the distances between corresponding tiles. This has the effect of ignoring large changes in some tiles brought about by camera or object movement rather than a change to a different scene. If the distance function d_2 between two adjacent frames, between frames 4 frames apart (d_4) and between 8 frames (d_8) peaks at the same frame, and the peaks are above a certain threshold then a cut boundary is called. Gradual transitions are detected through coinciding peaks of d_8 and d_{16} above another threshold. These two thresholds can be set via a test set of known videos in order to adapt them to the type of material. Rapidly changing MTV movies will need other thresholds than, say, the famously long scenes in a typical movie of the Russian director Tarkovsky. The video of Slide 15 (or directly from http://mmis.doc.ic.ac.uk/inaugural/anticipation-processed-2.mpg) visualises the process by plotting the distances d_2, d_4, d_8 and d_{16}. The top row indicates a cut via a | line and a gradual transition with a V sign. The movie itself is played at double speed, while the 3d texture false colour feature and the direct 3d colour histogram of the full frame are displayed for further illustration (though neither of these are directly used by our video shot boundary detection algorithm).

This analysis gives rise to creating *keyframes* from the video shots, which give a clear automated visual summary. Although there are sophisticated algorithms to identify the "best" frame from a shot (eg, by considering the blurriness of an image, the size of dominating objects, variance of parameters, etc) we use the 10th frame into a shot for simplicity and in order to avoid overlaid frames that tend to be present at the beginning or end of shots with gradual transitions. Slide 16 shows the keyframes from our automated process. Here is another example of this kind of visual summary in action in our news search engine:



Slide 16: Keyframes summary of *Anticipation*

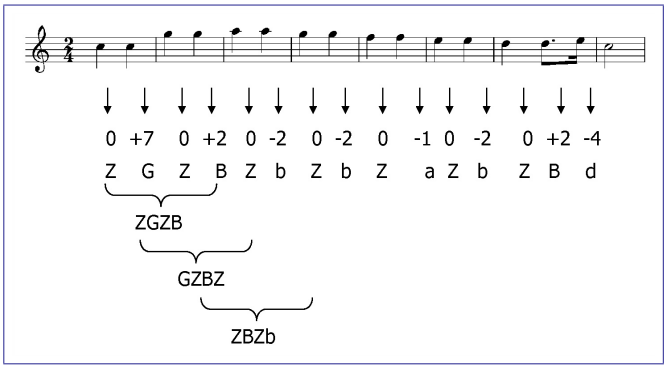
We record the 10 o'clock BBC news every day and process it automatically to segment it into cuts. Then the algorithm glues the shots together that belong to a story. This is early work, and a predecessor version by Marcus Pickering won the best computing student award in the UK in 2000, long before google and blinxtv launched their video search services. Let us search for Microsoft in the recent news. The first story in Slide 17 is about a failed attempt of Microsoft to buy Yahoo. You can clearly see a visual summary with Yahoo's logo, a summary of the recorded teletext, and automatically detected names of organisations, people, locations and dates. By virtue of automation these are bound to contain errors (though both Britain and the police actually feature in this story!). The next story, only present as visual summary in Slide 17, is about a release of a new game for the X-Box game console.

Slide 17: Anses: Automated News Summarisation and Extraction System

You will have noticed in this example that the search was a piggy-back search in the subtitles. We also try to provide as much useful information as possible. All is done automatically — there is no human involved. Following our mantra for added services we have created *shot detection*; *keyframe extraction*; *story segmentation*; *visual summary*; *text summary*; *organisation, people, location, date extraction*; and *full-text search*.

Our work in music retrieval has also also been based on reducing it to text retrieval. We

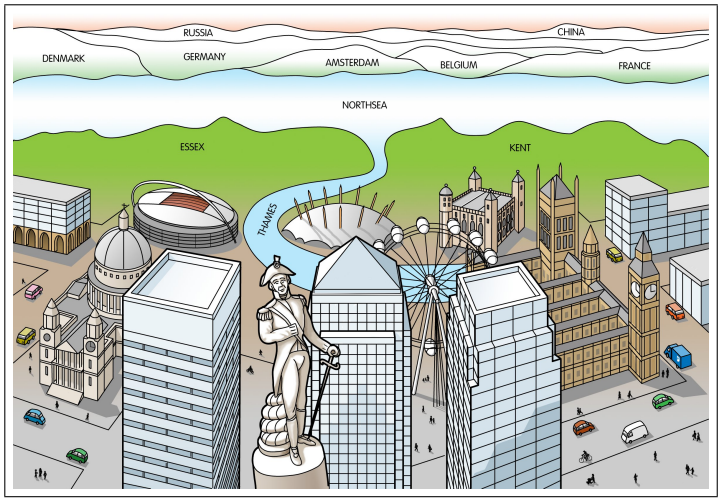
started with MIDI databases, which are a computerised version of the music score. The basic idea is to reduce the score to a text (Doraisamy and Rüger 2003).



Slide 18: Music Retrieval by Humming (click frame to play demo video)

The numbers are midi representations of the pitch of the score. We just record the differences of successive notes (as only a few gifted ones have the power of absolute pitch) and convert the difference to a letter. Zero is Z, 1 is upper case A, -1 is lowercase a, and so on. Then we glide a window over the music piece and record musical words of a certain length. This idea is not ours: it is from Stephen Downie, then at the University of Western Ontario. My former PhD student Shyamala Doraisamy started with this principle and extended it to both polyphonic music, where more than one note is present, and rhythm. She built a query by humming system that is based on the reduction of music to text followed by text search. She produced an impressive demonstration system that lays open the individual steps of her algorithm to show how it works: click on Slide 18 or go directly to http://mmis.doc.ic.ac.uk/inaugural/movie0008_audio.wmv.

Location, location, location. Geography is an important query element for any search. Why? The local area is vital to everyone, and this is why location is the prime context factor. There is a fantastic cartoon by Saul Steinberg from the cover of the 29 March 1976 edition of the *New Yorker*. This is my favourite cartoon ever and superbly expresses the prime role of location as context (rather than what others might argue New Yorkers' ignorance of the world map). It is called "View of the world from 9th Avenue". In this sketch you see the details of the 9th and 10th Avenue; then the Hudson river; behind that this map shows Jersey as a small area followed by the rest of the US, which takes up as much space as the area between 9th and 10th avenue in this map; then the Pacific Ocean as big as the Hudson River with a few indiscernible stretches of tiny land labeled as China, Japan and Russia.



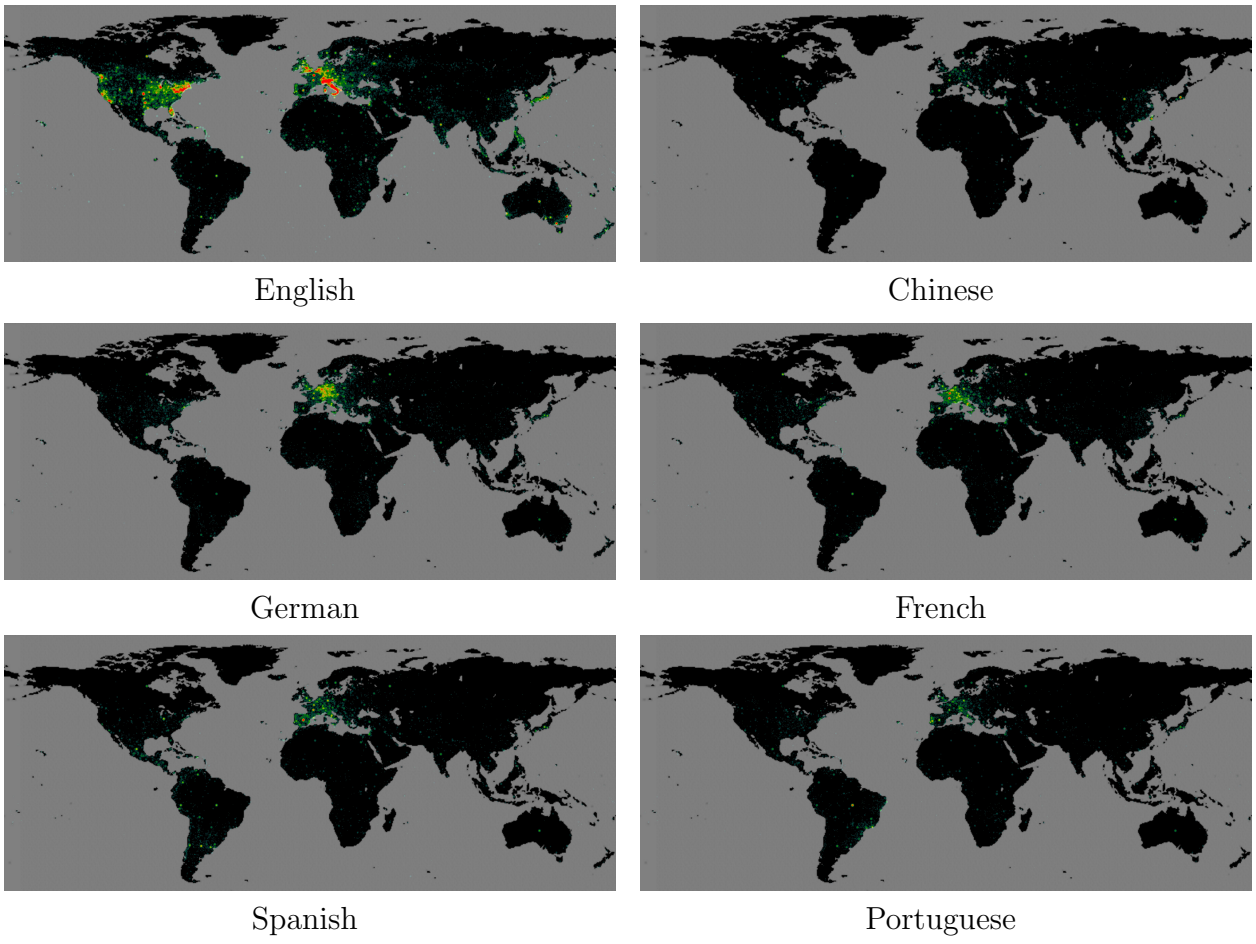
Slide 19: View of the world from Piccadilly Circus

Unfortunately, we were refused to buy the rights of this cartoon owing to the right holder's (Cartoon Bank) policy not to grant rights for re-publishing cover art on the web. You can find it on Cartoon Bank's own website, though, by searching for Steinberg world view.

Our genial lab artist Jon Linney created an analogous version for London's Piccadilly Circus, see Slide 19. I believe this is why location and geography are so important. One of my PhD students, Simon Overell, and I wanted to disambiguate placenames in free text. To do this we decided to build a model of how placenames occur in text

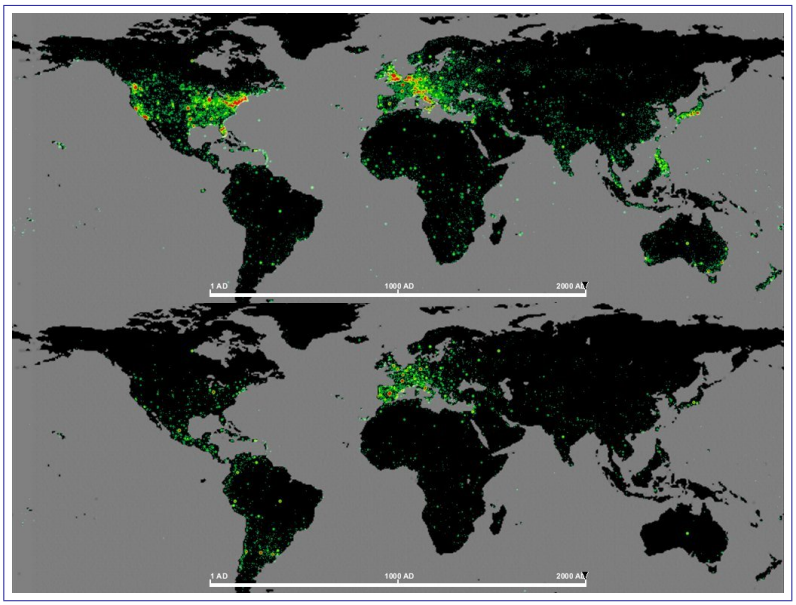
(Overell and Rüger 2007). One of the best resource to study this is wikipedia. Wikipedia is an on-line collaborative encyclopedia, where everyone can contribute articles about everything. When authors mention locations they usually link to an article about that location. This serves to disambiguate the location. Overell identified a way to find out which articles are about locations and also which specific latitude and longitude this would be. By looking at the 2m articles with 100m internal links in the English language wikipedia we obtained a huge corpus of how place names are used in the encyclopedia. The methods are language independent and we can do the same analysis for any of the 252 languages for which wikipedidae exist.

As a side effect of this pre-requisite for further research, we could visualise which locations are talked about how much in which language. It turns out that what people write about in wikipedia is strongly language dependent.



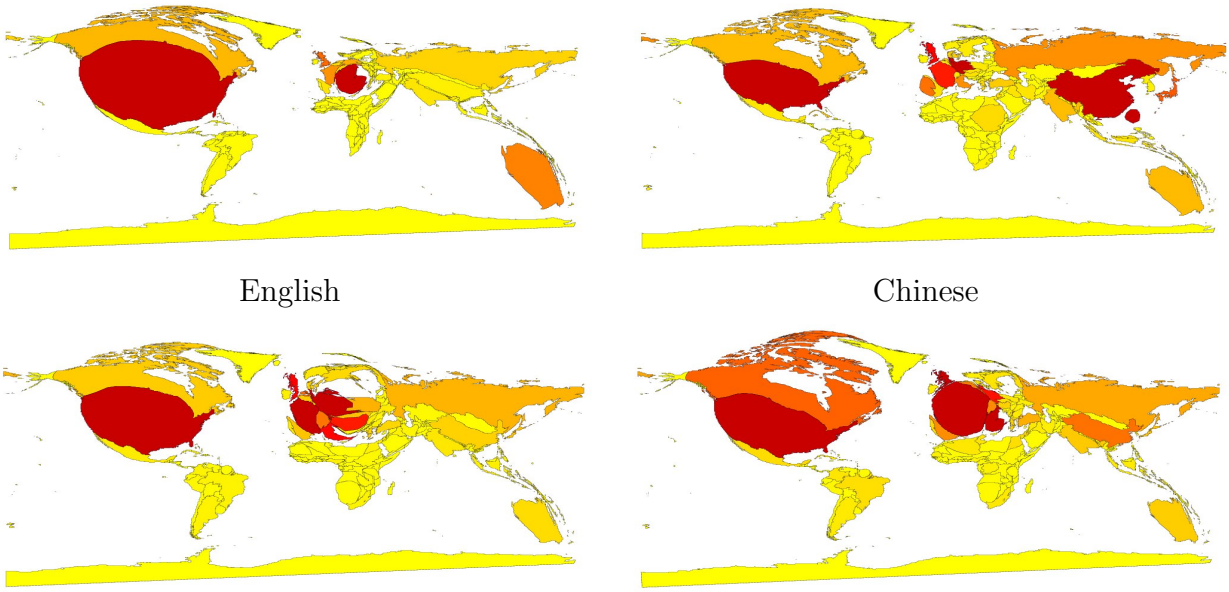
Slide 20: Heat maps of locations referred to in different languages

Here (Slide 20) are heat maps of locations that are talked about in English, Chinese, French, German, Portuguese and Spanish. Chinese has the most even distribution, as access in China is blocked from time to time and has been systematically blocked for more than a year. Hence, it is a high proportion of ex-patriates everywhere in the world, who contribute to wikipedia. To make it absolutely clear, this is about languages, not countries! Remember, too, that most languages are spoken in more than one country and that the majority of children in the world grow up with at least two languages.



Slide 21: Events from 1 CE to 2000 CE of the English (top) vs the Spanish (bottom) wikipedia (click frame for video)

Overell also created a video for extracted events from the year 1 CE to 2000 CE. Click at Slide 21 or directly at http://mmis.doc.ic.ac.uk/inaugural/events-en-sp-world.mpg to see for yourself how they compare in the English-language and Spanish-language wikipedia: Watch out for the discovery of America and how the Spanish and the English maps compare at this time. I find this absolutely fascinating! It is almost as if there were two completely different worlds. My hypothesis is that we all, not only the New Yorkers, behave, think and act like Saul Steinberg's cartoon suggests: Very local, indeed! Overell has visualised the density of references to places in cartograms, which are distorted world maps. The higher the density of references to locations in a particular country in comparison to what you would expect with respect to the population of that country, the bigger the exaggeration of the area of that country. Colour is another supporting visualisation means with yellow being an indicator for under-representative, orange for about right and red for over-representative number of references.

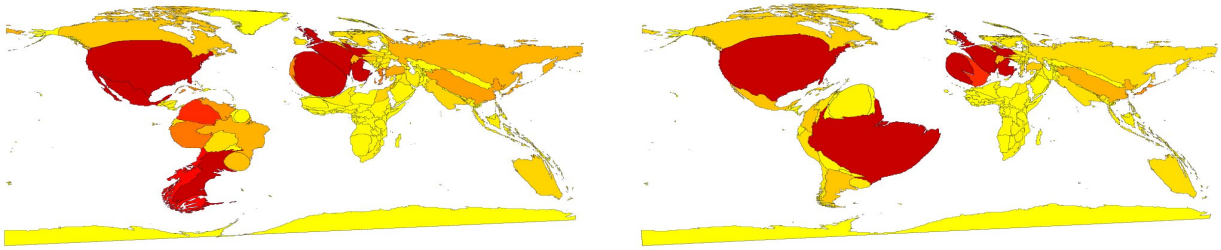


English Chinese

German French

Slide 22: Cartograms (distorted world maps)

The Spanish and Portuguese cartograms are my favourites:

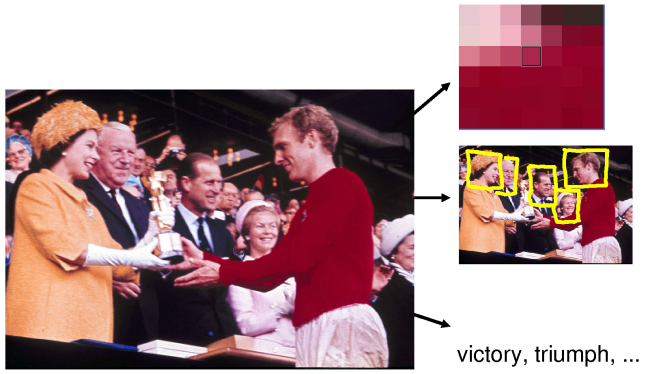


Spanish Portuguese

Slide 22: (cont'd) Cartograms

These maps convince me that we are all "little Steinbergs" be it with respect to location, our field of study, our expertise or otherwise. Hence, geography is one of the most important context factors for any kind of search engine and information provider.

The semantic gap and automated annotation. I will now come back to images and image processing. The biggest challenge here is the semantic gap. Look at this picture (Slide 23), for example. You may have seen it before.



Slide 23: The semantic gap

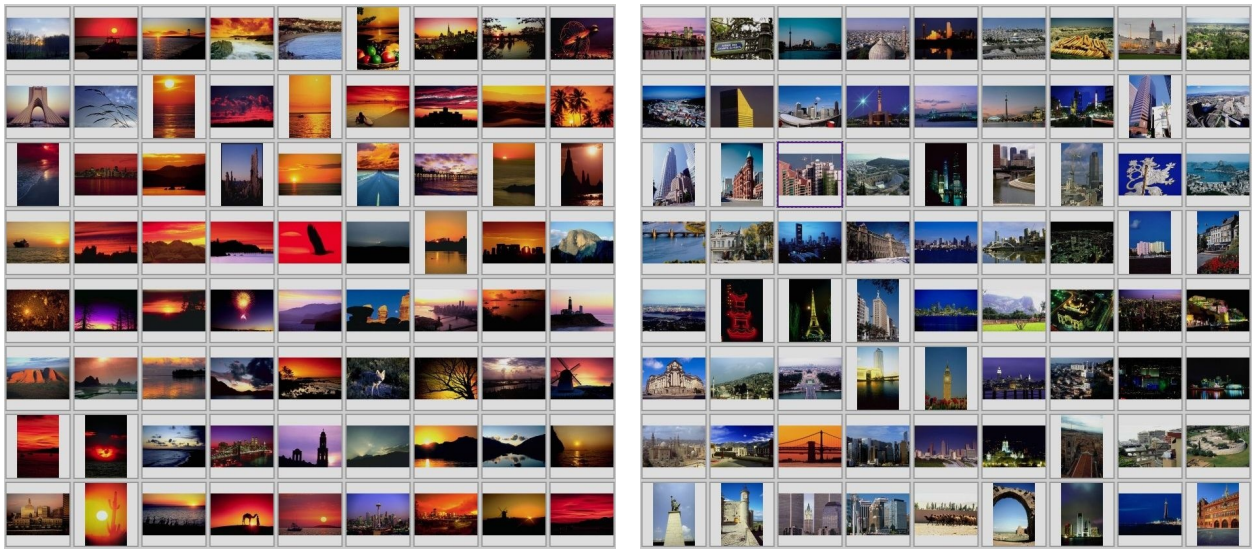
For a computer it signifies a number of pixels with a certain colour distribution. With state-of-the-art algorithms, one will find faces and certain object. But think again: this picture of the year 1966 shows the captain of a national team receiving a trophy from Her Royal Highness Queen Elizabeth II for winning a world tournament in just about the most important sports in the country! This image is really about glory, victory and triumph — if you support the English team, that is. For supporters of the West German team it probably signified misery, defeat and agony.

This discrepancy between low-level features and high-level is called the semantic gap — a hot problem in my field. One way to bridge it is to try to assign simple words automatically to images solely based on their pixels. One idea is to let computers automatically associate image features with certain words by the way of training data. We use Machine Learning with thousands of examples to learn automatically which features signify which label. Slide 25 shows randomly selected sunset images and randomly selected city images.

Slide 24: Automated annotation results in *water*, *buildings*, *city*, *sunset* and *aerial*

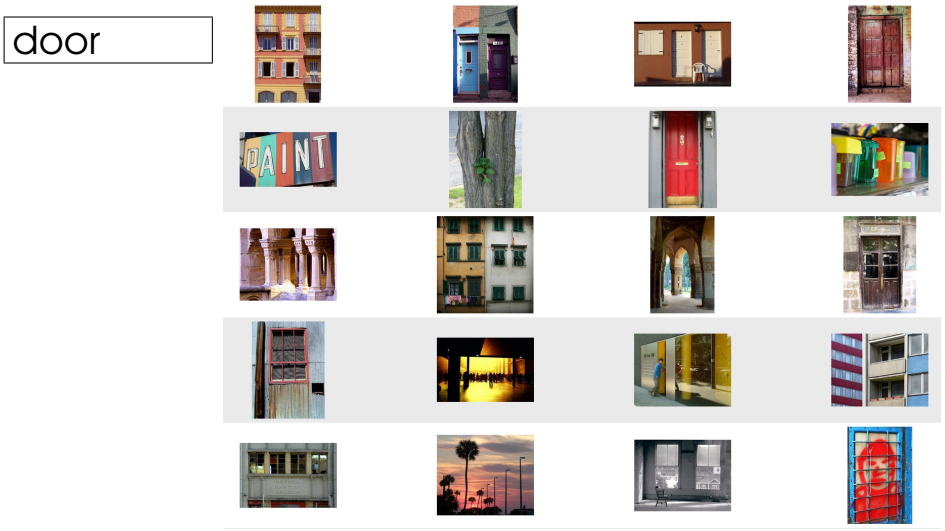
Automated algorithms build a model for the commonalities. These models can be used later for retrieval. Here is a result of one of our algo-

rithms (Naïve Bayes classification) on a new, previously unseen image. This algorithm was implemented by one of my former PhD students, Alexei Yavlinsky, and correctly predicts its labels! Yavlinsky went on to build even more successful algorithms (Yavlinsky et al 2005, Yavlinsky and Rüger 2007). For the specialists in the audience, these are non-parametric density estimators with specialised kernels that reflect the nature of the features. To date this is, to the best of my knowledge, still the most precise algorithms around.

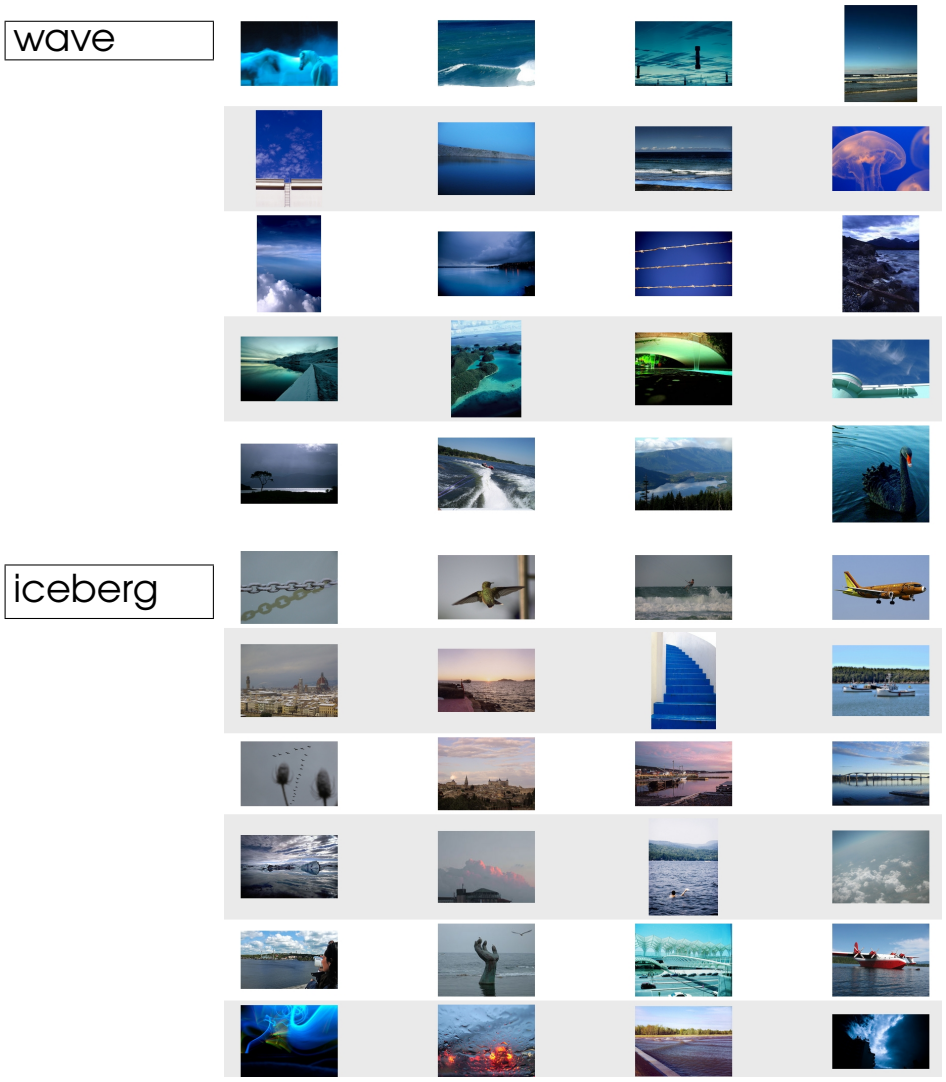


Slide 25: Machine learning training samples for *sunset* and *city* images

Yavlinsky built a corresponding search engine, where one could search for flickr images using these detected terms. Another of my PhD students, João Magalhães, has improved the speed and flexibility of solutions to this problem with alternative machine learning algorithms placing himself in the top for usability (Magalhães and Rüger 2007). These algorithms all make errors as you can expect from fully automated systems. Here are are screenshots from a search engine behold that was developed by Yavlinsky during his PhD time:



Slide 26: The good, ...



Slide 26: (cont'd) ..., the bad and the ugly: images that were automatically annotated

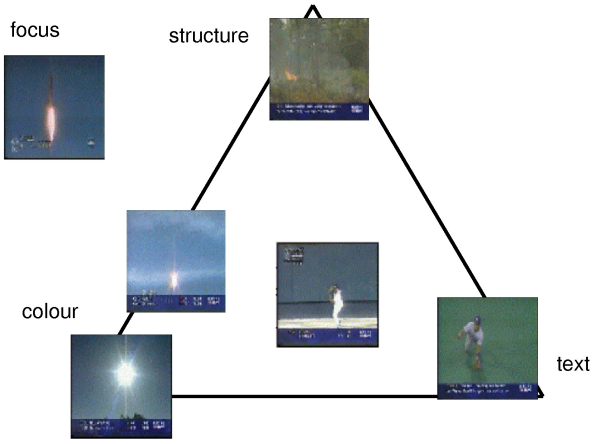
Clearly, not all words are predicted correctly, and currently one of my PhD students, Ainhoa Llorente, tries out ways to introduce world knowledge into the process, eg, that stairs and icebergs normally don't go together.

So far we have started bridging the semantic gap a little towards words that have a visual correspondence (sunsets as opposed to victory).

Browsing, the Cinderella of Information Retrieval. There is one last important aspect of multimedia resource discovery that I have not yet mentioned. This is the process of browsing, the Cinderella of Information Retrieval. Browsing follows paths between objects. One way to make browsing fun and interesting is to create a network between all objects in a media database, so that any two objects can be reached by following short paths.

We all have heard the phrase *six degrees of separation*. This goes back to an experiment by Stanley Milgram in 1967. He asked 160 randomly selected people from Wichita, Kansas, and Ohama, Nebraska, to forward a message to a particular stockbroker in Boston (presumably these two cities were just Steinberg's version of "just out there"). They were only allowed to send the message through their social network of people they were on 1st name basis with

(today this would probably mean "people you can pull a favour from"). Milgram followed these paths and found that, of all paths that succeeded, there were on average 5.5 links involved. In today's world this means if you wanted to be introduced to super model Heidi Klum or actor George Clooney, as the case may be, you could probably ask in your social network someone who presumably is a bit closer to them to ask who they think is a bit closer and so on. If you are lucky this process will take around 6 steps.



Slide 27: NN^k network construction

Using this principle of a small degree of separation one can try to construct a network in media objects. Slide 27 shows how the (arbitrary) focus image can be connect within the network. The construction is fairly simply: link this image to all images that are closest in some respect, be it colour, shape, texture, edge usage, ... Not only single features can be used with this principle: *Any* weighted convex combination of distances of constituent features gives rise to a closest image that can be added to the set of images that are closest in *some* respect. Doing this right will connect the database with 3 degrees of separation.

Slide 28 visualises an example information need that we will try solving with browsing alone — without any text searches or example images! This particular information has been taken from the TRECVid evaluation conference series. The task is *to find video shots from behind the pitcher in a baseball game as he throws a ball that the batter swings at.*

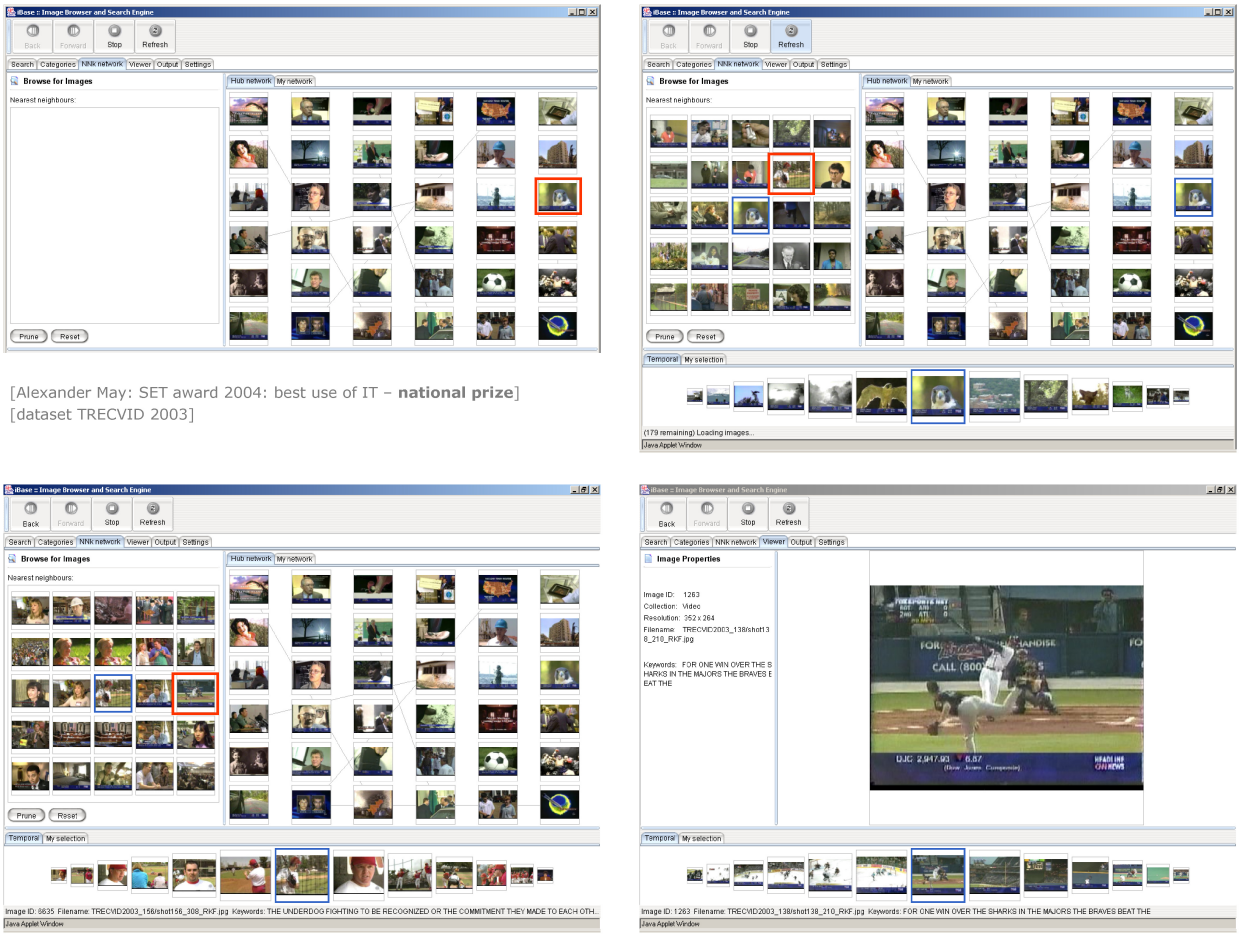
Slide 28: TRECVid topic 102

We try to solve this with a visual browsing engine that also won a national prize for Alexander May, an undergraduate student at Imperial College London. This is shown in the upper left part of Slide 29. It is initialised with random well-connected starting points. We are looking for a sports field — something green. So we select the falcon over here. Clicking on the falcon will show his neighbours in the pre-computed network.

There is an image that looks a bit more like a sports field. Clicking on this will in turn show its neighbors (Slide 29 upper right).

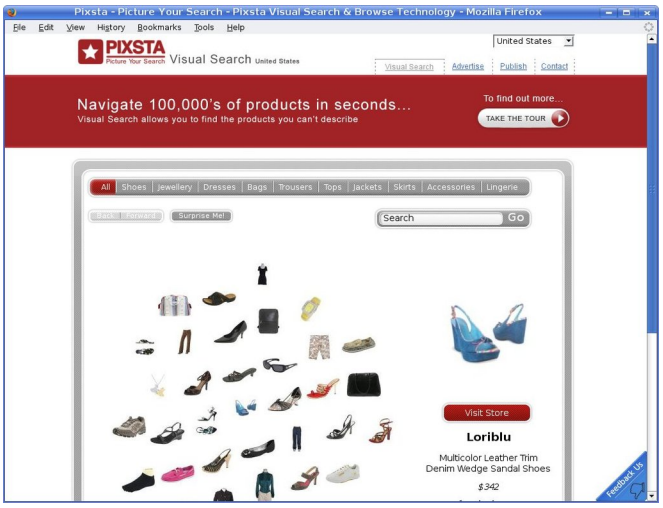
Examining the neighbors here (Slide 29 lower left) we see that we stroke lucky with one particular image enlarged in Slide 29 lower right.

Summarising, it was three clicks browsing, our imagination and a clever arrangement of the objects that got us the results!



[Alexander May: SET award 2004: best use of IT – **national prize**]
[dataset TRECVID 2003]

Slide 29: Browsing for shots "from behind the pitcher ..." with 3 clicks



Daniel Heesch, one of my former PhD students, who masterminded this process, has in the meantime co-founded a company called Pixsta that offers browsable electronic retail catalogues based on this idea.

Here (Slide 30) you see a screenshot of his company's home page. Ryan, to answer your question about the Argos catalogue: Good guess, but no, it is not me, rather one of my former PhD students, who does that!

Slide 30: Pixsta Ltd uses lateral browsing

This concludes the formal part of my inaugural lecture, but please allow me to thank all those who have made all this research possible.

Acknowledgements. Undoubtedly, one of the joys of an academic post is travelling: not only the physical travel to meet colleagues to argue, discuss and work with them — I have worked in Berlin, London, Tokyo, Sydney, Hong Kong, Hamilton, Bayreuth, Grenoble and Milton Keynes — but also the mental travel in the world of concepts, paradigms and knowledge. I have travelled from a world of theoretical physics to a world of applications in multimedia retrieval, and later from one of the most exclusive universities, Imperial College London, to one of the most inclusive universities, The Open University, UK, whose motto is to be open to people, methods, places and ideas, something that I very much support.

Slide 31: Star wars storm troopers and Darth Maul (click image for the credits movie)

As an aside, when I informed my then 7-year-old son about the imminent move, he was very, very happy about this. When he saw my puzzled expression he explained that — as a Star Wars devotee — he could never understand why I had worked for the Empire in the first place (presumably thinking I would teach Imperial stormtroopers how to quash Jedi rebels).

During my many travels in life I have met and worked with a large number of people without who I could never have undertaken all these journeys, be it because they inspired me, because they collaborated with me, because they supported me or because they kept the more crazy of my ideas in check, ... It is impossible to list them all and impossible too to adequately recognise and value their input. There are, of course, my immediate family, my mother and my late father, Irene and Josef, who have given me the most fantastic childhood, my wife Theresa Erhard, and my son Ryan, my brothers and their families, my teachers and supervisors at university, my co-students, and my own staff and students in my group, co-authors of scientific papers and grant proposals and colleagues who worked with me on joint projects.

I thank them all.

Episode 4 — Credits

Cuneiform script tablet, the Kirkor Minassian collection in the Library of Congress, ca 24th century BCE

Colossus, Public Record Office, London

Vannevar Bush portrait, United States Library of Congress's Prints and Photographs Division, digital ID cph.3a37339

Memex design, redrawn by Jon Linney based on http://www.kerryr.net/pioneers/gallery/ns_bush8.htm (2 June 2008) from a drawing originally found in Life Magazine, 19 November 1945, p123.

Text index, DK Eyewitness Travel Guides, New Zealand, Dorling Kindersley, 2001 — reprinted 2002

Peace Pagoda, Stefan Rüger, 22 July 2007

Snap.Send.Get^{TM}, with kind permission from SNAPTELL Inc

Medical images, Image CLEF 2004 corpus

Theresa and columns, Aarhus Art Museum, Stefan Rüger, May 1996

3d colour histogram visualisation by Anuj Kumar, May 2008

Textures, Stefan Rüger, June 2008

Texture computation programme, Peter Howarth, 2004

Anticipation, a film by Vlad Tanasescu with Claudio Baldassarre, Silvia Cassese, Sohan Jheeta and the voice of Samuel Spycher; lights and production by Chris Valentine; music: Johannes Brahms, Op. 45, Ein deutsches Requiem "Denn alles Fleisch, es ist wie Gras", the Holden Consort Orchestra and Choir; media: Statue of Liberty and WTC pic 1 from the National Park Service at http://www.nps.gov, Statue of Liberty and WTC pic 2 from Marvin at Flickr.com; objects by Ernest von Rosen at http://www.amgmedia.com

Visualisation of shot boundary detection by Marcus Pickering, May 2008

Keyframe computation with hive2, a shot boundary detection programme by Marcus Pickering, 2000, modified by Eric Fernandez, 2007

ANSES, originally by Marcus Pickering, 2000, modified by Lawrence Wong, 2004

Music retrieval demo, Music Retrieval system interface and video production, Shyamala Doraisamy and Kok Huai Meian, University Putra Malaysia; Bach's Fugue No.1 in C major BWV 846 and other MIDI files from http://www.classicalarchives.com; voice: Kok Huai Meian; humming: Shyamala Doraisamy

Queen and Booby Moore, © Associated Press/Empics, used with permission

New Yorker cover of 29 March 1976, © New Yorker/Cartoon Bank, used with permission during lecture — not shown on the internet or replay of lecture

View of the world from Piccadilly Circus with kind permission from Jon Linney free after the idea of above *New Yorker* cover, used for the post-lecture notes

Wikipedia logo, © and registered trademark of Wikimedia Foundation Inc

Heat world map video and images, Simon Overell using wikipedia and NASA maps from Blue Marble: Next Generation, NASA's Earth Observatory

Cartograms, Simon Overell using the application MAPresso from http://www.mapresso.com

NY city image from Corel Gallery 380,000 — royalty free images

Behold, by Alexei Yavlinsky, screenshots from http://photo.beholdsearch.com, 19 July 2007, now http://www.behold.cc

Rocket images and baseball images from the TRECVid 2003 corpus

uBase written 2004 by Alexander May with back-end components by Daniel Heesch, Peter Howarth, Marcus Pickering, Alexei Yavlinsky and, later, Paul Browne

Pixsta screenshot (http://www.pixsta.com), 30 May 2008, with kind permission from Pixsta Ltd

Stormtroopers (Lego figures), Stefan Rüger, June 2008

Episode 5 — References and further reading

A Dalby: The Sumerian catalogs, *Journal of library history*, 21(3), Summer 1986

S Doraisamy and S Rüger: Robust polyphonic music retrieval with *n*-grams. *Journal of Intelligent Information Systems*, 21(1), pp 53–70, 2003

S Doraisamy: *Polyphonic music retrieval: the n-gram approach.* PhD Thesis, Imperial College London, 2004.

D Heesch and S Rüger: *NN^k networks for content-based image retrieval.* European conf on Information Retrieval (ECIR, Sunderland, UK), Springer LNCS 2997, pp 253–266, Apr 2004

D Heesch and S Rüger: *Image browsing: semantic analysis of NN^k networks.* Int'l conf on Image and Video Retrieval (CIVR, Singapore), Springer LNCS 3568, pp 609–618, Jul 2005

D Heesch, P Howarth, J Magalhães, A May, M Pickering, A Yavlinsky and S Rüger: *Video retrieval using search and browsing.* Notebook of the TREC video retrieval evaluation (TRECVid, Gaithersburg, MD), Nov 2004

D Heesch, A Yavlinsky and S Rüger: *NN^k networks and automated annotation for browsing large image collections from the world wide web.* Proc of the ACM conf on Multimedia (ACM MM, Santa Barbara, CA), Oct 2006

D Heesch: *The NN^k technique for image searching and browsing.* PhD Thesis, Imperial College London, Dec 2005

P Howarth and S Rüger: Robust texture features for still-image retrieval. *IEE Proc on Vision, Image and Signal Processing*, 152(6), pp 868–874, 2005

P Howarth, A Yavlinsky, D Heesch and S Rüger: *Medical image retrieval using texture, locality and colour.* Springer LNCS 3491, pp 740–749, 2005

P Howarth: *Discovering images: features, similarities and subspaces.* PhD Thesis, Imperial College London, June 2007.

J Magalhães and S Rüger: *Information-theoretic semantic multimedia indexing.* Int'l ACM conf on Image and Video Retrieval (CIVR, Amsterdam, The Netherlands), Jul 2007, *best paper award*

S Overell and S Rüger: *Geographic co-occurrence as a tool for GIR.* CIKM Workshop on Geographic Information Retrieval (CIKM, Lisbon, Portugal), pp 71–76, Nov 2007

M Pickering, L Wong and S Rüger: *ANSES: Summarisation of news video.* Int'l conf on Image and Video Retrieval (CIVR, Urbana-Champaign, IL), Springer LNCS 2728, pp 425–434, Jul 2003

M Pickering: *Video retrieval and summarisation.* PhD Thesis, Imperial College London, July 2004.

A Yavlinsky and S Rüger: *Efficient re-indexing of automatically annotated image collections using keyword combination.* Proc Multimedia Content Analysis, Management and Retrieval (SPIE, San Jose, CA), Jan 2007

A Yavlinsky, E Schofield and S Rüger: *Automated image annotation using global features and robust nonparametric density estimation.* Int'l conf on Image and Video Retrieval (CIVR, Singapore), Springer LNCS 3568, pp 507–517, Jul 2005

A Yavlinsky: *Image indexing and retrieval using automated annotation.* PhD Thesis, Imperial College London, Sept 2007

Episode 2 — Physics and Computer Science studies

Klaus Fredenhagen . . . . . coauthor, supervisor
Matthias Gaberdiel . . . . . coauthor
Thomas Kessler . . . . . coauthor
Klaus Obermayer . . . . . supervisor
Arnfried Ossen . . . . . coauthor (7x)
Anton Weinberger . . . . . coauthor, Dipl-Inform
Sebastian Wittchen . . . . . coauthor, Dipl-Inform
Wolf-Dieter Woidt . . . . . coauthor

Episode 3 — Academic career

Peter Kwok Tat Au . . . . . coauthor
David Bainbridge . . . . . coauthor, MMDL, visitor, host
Cristi Barladeanu . . . . . coauthor
Eric Blanco . . . . . coauthor
Peter Bruza . . . . . coauthor (2x)
David Bull . . . . . coauthor
Paul Cairns . . . . . coauthor (2x), MMDL
Bejal Chawda . . . . . coauthor, MSc
Richard Cooper . . . . . coauthor, MSc
Brock Craft . . . . . coauthor
Greg Crane . . . . . CHLT Co-I
John Darlington . . . . . coauthor
John Eakins . . . . . MMKM steering group
Marc Eisenstadt . . . . . coauthor (2x)
Peter Enser . . . . . MMKM steering group
David Feng . . . . . host
Eric Fernandez . . . . . MSc
Jodie Forbes-Millott . . . . . coauthor
Mickaël Gardoni . . . . . coauthor (2x), visitor, host
Susan Gauch . . . . . coauthor
Julien Gevrey . . . . . coauthor, MSc
Arnab Ghoshal . . . . . coauthor
Duncan Gillies . . . . . coauthor
Alexandre Gonçalves . . . . . coauthor
Yike Guo . . . . . coauthor (3x)
Alex Hauptmann . . . . . coauthor
J Hoare . . . . . coauthor
Ian Horrocks . . . . . MMKM steering group
Ebroul Izquierdo . . . . . MMKM steering group
Rob Iliffe . . . . . CHLT Co-I
Dolores Iorizzo . . . . . CHLT Co-I
Rui Jesus . . . . . coauthor, visitor
Sanjeev Khudanphur . . . . . coauthor
Mounia Lalmas . . . . . coauthor (2x), co-organiser ECIR 2006, Renaissance Co-I
Paul Lewis . . . . . MMKM steering group

Frederic Fol Leymarie . . . . . MMKM steering group
Andrew MacFarlane . . . . . coauthor, co-organiser ECIR 2006
R Manmatha . . . . . coauthor
Yosi Mass . . . . . coauthor
Enrico Motta . . . . . coauthor (2x)
Robert O'Callaghan . . . . . coauthor
Roberto Pachero . . . . . coauthor
Maja Pantic . . . . . MMKM steering
Jeffrey Rydberg-Cox . . . . . coauthor (2x), CHLT Co-I
Mark Sanderson . . . . . coauthor, MMKM steering
Shalini Sewraz . . . . . coauthor (2x), MSc
Heng Tao Shen . . . . . coauthor
David Sinclair . . . . . coauthor
Janjao Sutiwaraphun . . . . . coauthor
Anastasios Tombros . . . . . coauthor
Theodora Tsikrika . . . . . coauthor (2x)
Vicoria Uren . . . . . coauthor (2x)
Roelof van Zwol . . . . . coauthor
Keith van Rijsbergen . . . . . MMKM steering, Renaissance Co-I
Lara Vetter . . . . . coauthor (2x)
Thomas von Schröter . . . . . coauthor
Lawrence Wong . . . . . coauthor, MEng
Li-Qun Xu . . . . . coauthor (2x), MMDL
Ian Witten . . . . . host
Zdenek Zdrahal . . . . . PAHROS, Co-I
Xiaofang Zhou . . . . . coauthor

Episode 6 — Team members, past & present


Paul Browne . . . . . coauthor (2x), MMDL, postdoc

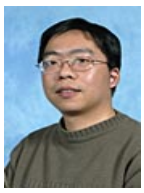

Matthew Carey . . . . . coauthor (2x), CHTL, MSc


Shyamala Doraisamy . . . . . coauthor (7x), PhD


Daniel Heesch . . . . . coauthor (23x), CHLT RA, MSc, PhD


Peter Howarth . . . . . coauthor (8x), PhD


Rui Hu . . . . . coauthor (2x), PhD, MMKM


Qiang Huang . . . . . coauthor (3x)


Zi Huang . . . . . coauthor (2x), PHAROS postdoc


Partha Lal . . . . . coauthor, MSc


Hai-Ming Liu . . . . . coauthor (2x), PhD


Ainhoa Llorente Coto . . . . . PhD


João Magalhães . . . . . coauthor (13x), PhD

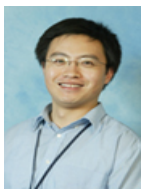

Alexander May . . . . . coauthor, MEng


Simon Overell . . . . . coauthor (8x), PhD


Marcus Pickering . . . . . coauthor (11x), PhD

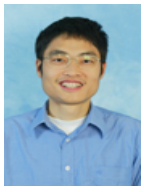

Adam Rae . . . . . coauthor, PhD


Edward Schofield . . . . . coauthor, PhD

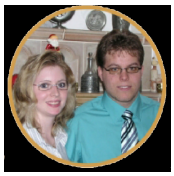

Dawei Song . . . . . coauthor (10x), PHAROS Co-I, Renaissance PI

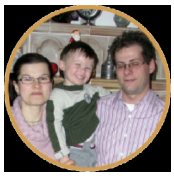

Alexei Yavlinsky . . . . . coauthor (15x), PhD, MMKM


Zrdan Zagorac . . . . . PHAROS research developer


Jianhan Zhu . . . . . coauthor (4x), PHAROS postdoc

Episode 1 — family

 Alexander and Monika Rüger

 Thomas, Kathrin and Jonas Rüger

 Dieter and Gabi Rüger

 Stefanie Rüger

 Robert and Bernadette Rüger

  Bernhard Rüger, Annette Kober

  Josef and Irene Rüger

 Theresa Erhard, Ryan & Stefan Rüger