Centre for
Data Analytics

Insight

(Digital)
Cultural Heritage: From Preservation to Access to Promotion to Where ?

Alan F. Smeaton
Dublin City University

Presentation Overview

- Apologies to Thomas C. Südhof, “Mechanisms of Neurotransmitter Release”, Nobel Prize winner Physiology/Medicine 2013
- Digital Cultural Heritage … 20 years, to now
- Capturing intangible culture
- Preservation – Accessibility – Access – Promotion
- Looming hurdle in promoting digital CH

  … content-based access
- Content access for CH digital imagery

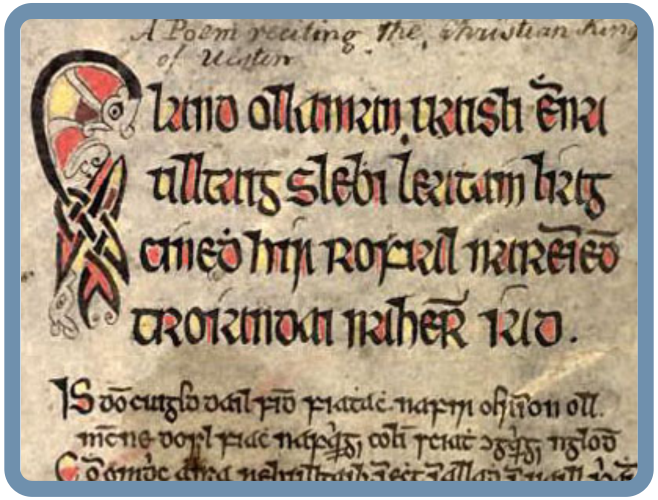
Digital Cultural Heritage Started

- Digitisation became cost-effective 20 years ago
- Clamor for CH digitisation, national, EU, philantrophic

, National Library of Ireland, MS. 700, f. [39] v
tional Library of Ireland, 2002

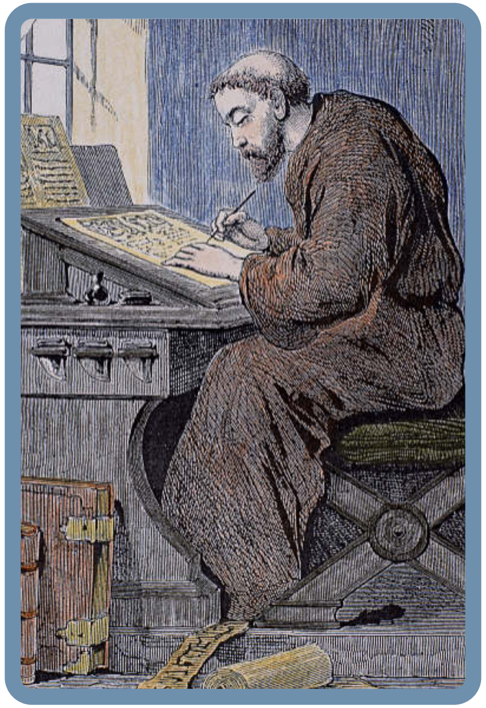
ISOS

- **Irish Script On Screen**
  - DIAS project with DCU assistance
  - Make digital images of ancient Irish manuscripts available in an online Digital Library
    - Small (150-500K) JPEG images
    - Large (1.5-5MB) JPEG images
  - http://www.isos.dias.ie/
  - Metadata includes manuscript details and page numbering



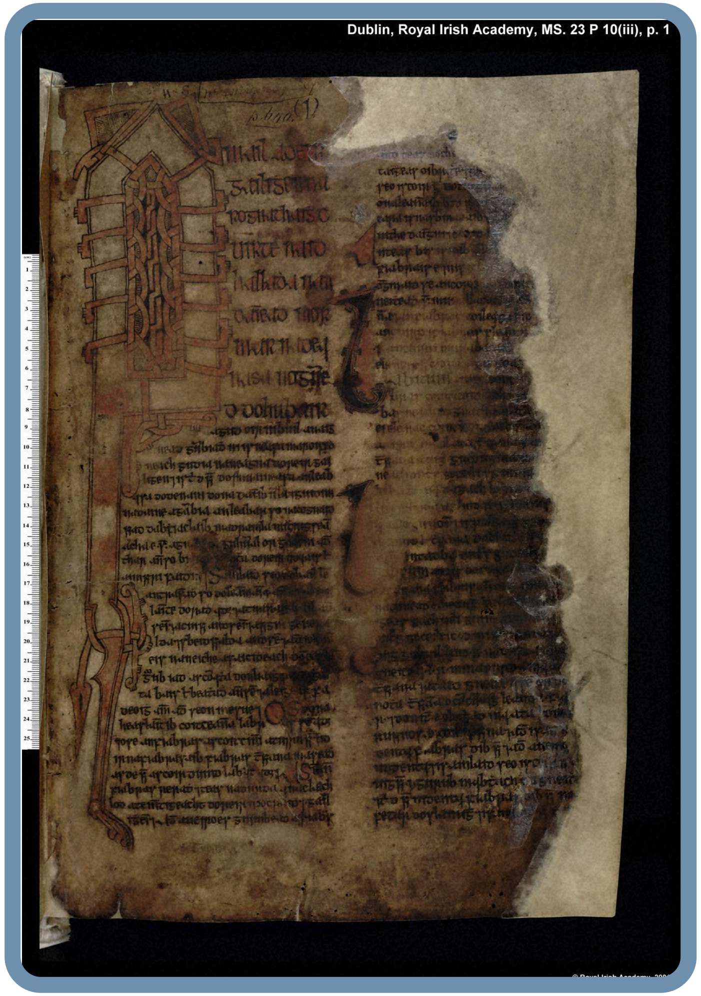
- **The collection contains manuscript pages from manuscripts from collections from universities, museums and Abbeys**
- **Helps to make ancient Irish manuscripts available to general public as well as to Celtic scholars**

ISOS Important Features



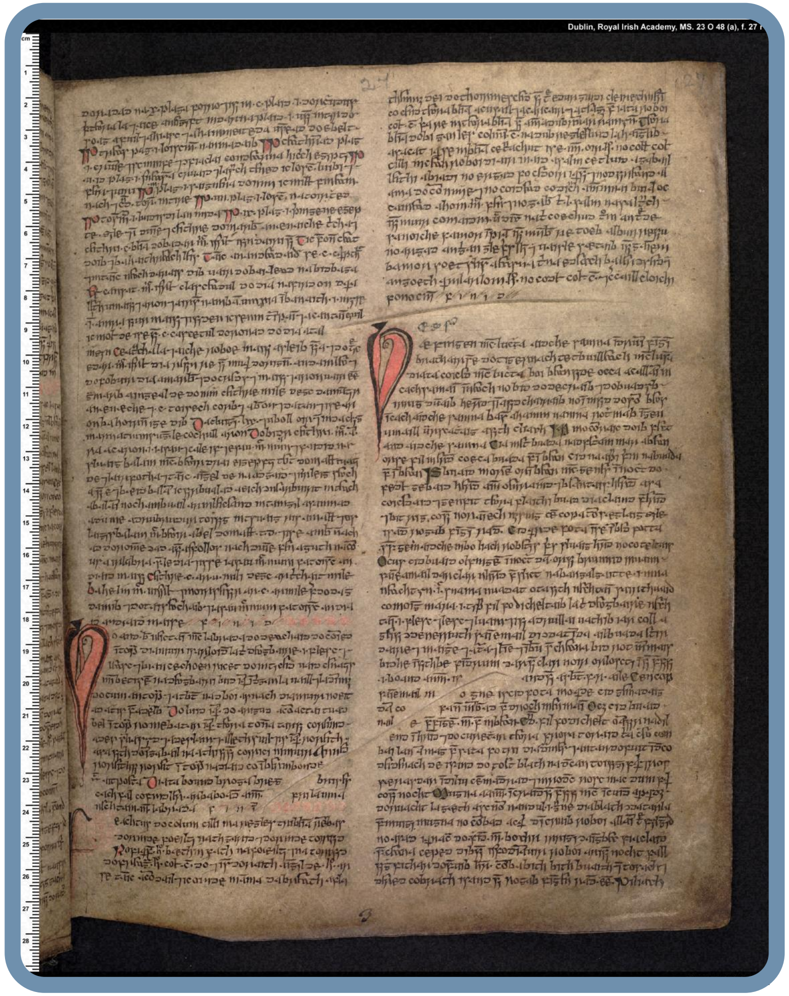
- **Strict Editorial control of the objects in the library**
  - Not an upload process
  - **An expensive digitisation process means that careful planning is required before any manuscript is included.**
- **No need for people to travel to access the library and the ISOS digital library is available a**
- **Anyone can access the collection**
  - Though large JPEG images are reserved for registered library users
- **ISOS helps to digitally preserve degrading manuscripts**

ISOS Manuscripts



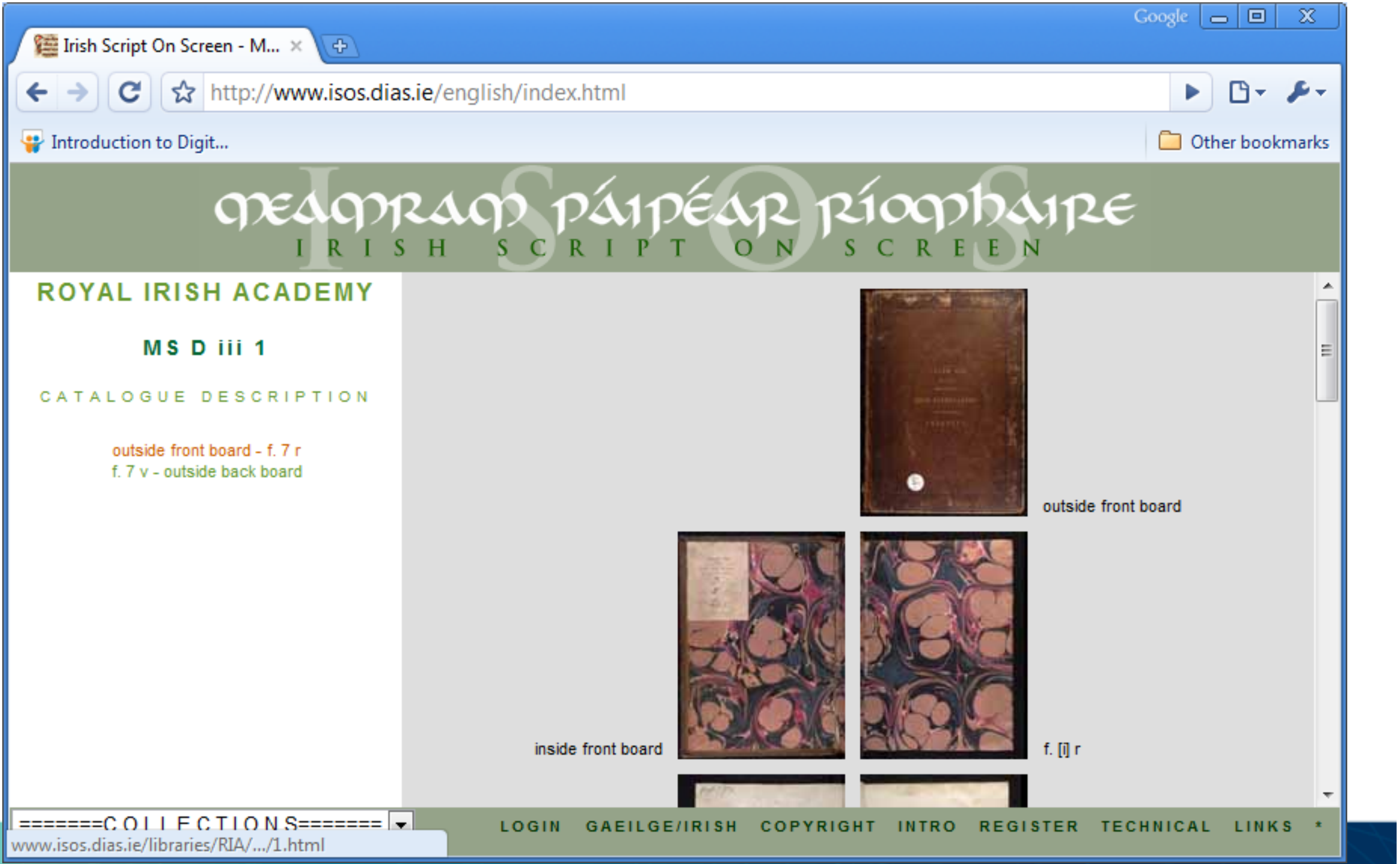
Dublin, Royal Irish Academy, MS. 23 P 10(iii), p. 1

© Royal Irish Academy, 200



Dublin, Royal Irish Academy, MS. 23 O 48 (a), f. 27 r

ISOS Website



Irish Script On Screen - M...

http://www.isos.dias.ie/english/index.html

Introduction to Digit...

Other bookmarks

MEAMRAM PÁIPÉAR RÍOMHAIRE

IRISH SCRIPT ON SCREEN

ROYAL IRISH ACADEMY

MS D iii 1

CATALOGUE DESCRIPTION

outside front board - f. 7 r
f. 7 v - outside back board

outside front board

inside front board

f. [i] r

=======COLLECTIONS=======

LOGIN GAEILGE/IRISH COPYRIGHT INTRO REGISTER TECHNICAL LINKS *

www.isos.dias.ie/libraries/RIA/.../1.html

ISOS Purpose

- So the aims were PRESERVATION and ACCESSIBILITY (not access or promotion !)

www.isos.dias.ie

- Content-based Access non-existent though there was some work on using word shapes as a basis for retrieval

- Group similarly-shaped characters using a reduced alphabet, based on considerations of appearance of letters

Word Shape Mapping

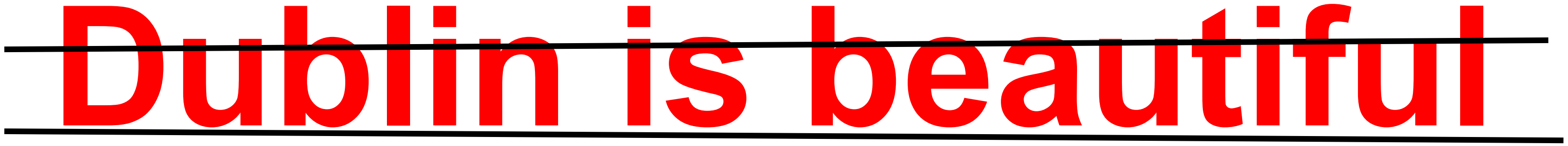
A← A-Z b d f h k l t

x ← a c e m n o r s u v w x z

g ← g p q y

i ← i

j ← j



Dublin is beautiful

- … maps to “AxgxxxAA ix AxxxAiAxA”

Word Shape Mapping

Based on a lexicon of 318,636 correct word forms, many share the same WST

xxxAxxx = numbers, nuclear, murders, and 281 others

xxxA = card, rush, cost, and 204 others

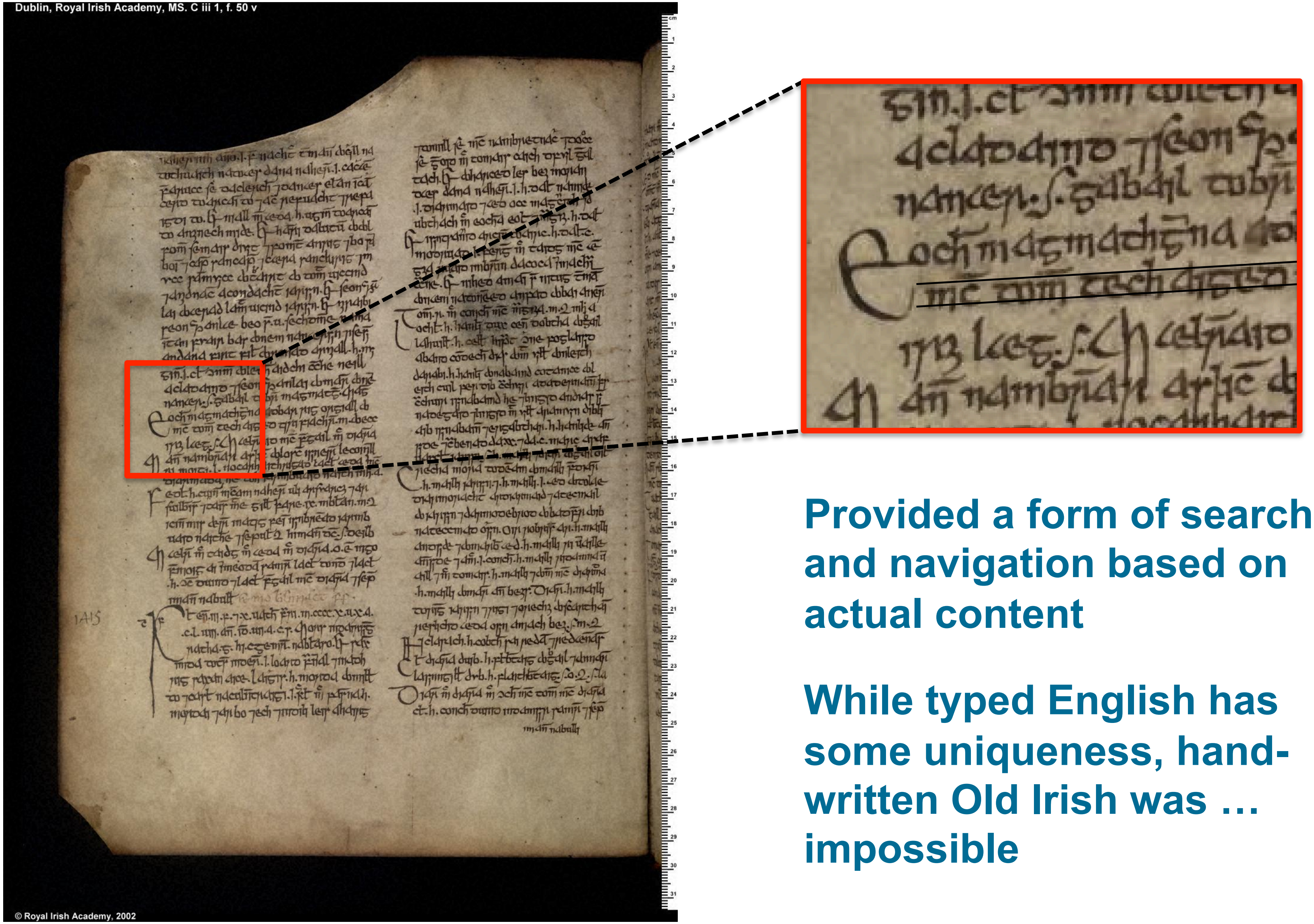
but

xxAxixxxAAx = retrievable, untrimmable

xxAxixxx = retrieve, cadmiums, calcines, entwines, matrices, retainer, sobrinus

while the following are all unique

automobiles, Environmental, contribute, criminals, military, ignored, origin, vitamin



Dublin, Royal Irish Academy, MS. C iii 1, f. 50 v

© Royal Irish Academy, 2002

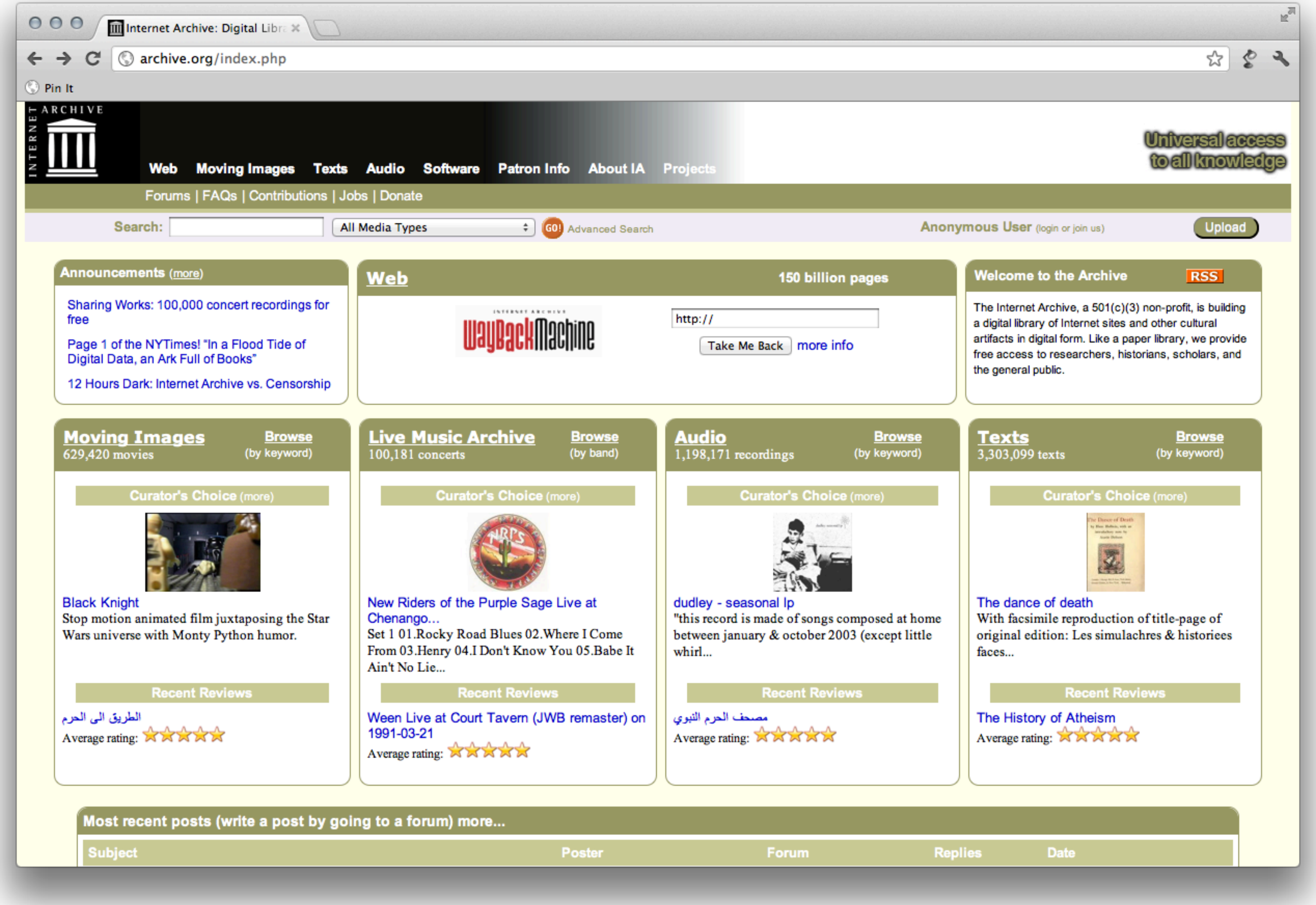
Provided a form of search and navigation based on actual content

While typed English has some uniqueness, hand-written Old Irish was … impossible

Cultural Heritage

- Much early digital CH was Preservation and Accessibility .. what about the internet (not the same as social media) ?
- The Internet Archive … archive.org … non-profit digital library, free public access … 200 employees scanning books, \$10m per annum,
- 10 petabytes 3 years ago = 10,000,000,000,000,000 B = 10^{16} bytes = 400 Billion web pages
- Slow to retrieve because roboticised storage!

Archive.org



Announcements (more)

Sharing Works: 100,000 concert recordings for free

Page 1 of the NYTimes! "In a Flood Tide of Digital Data, an Ark Full of Books"

12 Hours Dark: Internet Archive vs. Censorship

Web 150 billion pages

Welcome to the Archive RSS

The Internet Archive, a 501(c)(3) non-profit, is building a digital library of Internet sites and other cultural artifacts in digital form. Like a paper library, we provide free access to researchers, historians, scholars, and the general public.

Moving Images 629,420 movies — Browse (by keyword)

Curator's Choice (more)

Black Knight
Stop motion animated film juxtaposing the Star Wars universe with Monty Python humor.

Recent Reviews

الطريق الى الحرم
Average rating: ★★★★★

Live Music Archive 100,181 concerts — Browse (by band)

Curator's Choice (more)

New Riders of the Purple Sage Live at Chenango...
Set 1 01.Rocky Road Blues 02.Where I Come From 03.Henry 04.I Don't Know You 05.Babe It Ain't No Lie...

Recent Reviews

Ween Live at Court Tavern (JWB remaster) on 1991-03-21
Average rating: ★★★★★

Audio 1,198,171 recordings — Browse (by keyword)

Curator's Choice (more)

dudley - seasonal lp
"this record is made of songs composed at home between january & october 2003 (except little whirl...

Recent Reviews

مصحف الحرم النبوي
Average rating: ★★★★★

Texts 3,303,099 texts — Browse (by keyword)

Curator's Choice (more)

The dance of death
With facsimile reproduction of title-page of original edition: Les simulachres & historiees faces...

Recent Reviews

The History of Atheism
Average rating: ★★★★★

Most recent posts (write a post by going to a forum) more...

Subject	Poster	Forum	Replies	Date

Sample Websites

Name	Website	Wayback Machine Records	Total links	Rank(Global & Ireland)
Agriland	agriland.ie	40(2012-2015)	377125	292,928/2,061
Labour	labour.ie	504(1998-2014)	1063988	1,060,560/6,273
Fianna Fail	fiannafail.ie	406(1998-2014)	735891	1,408,305/6,663
Fine Gael	finegael.ie	293(1998-2014)	882846	807,786/4,330
Sinn Fein	sinnfein.ie	577(1998-2012)	1884479	615,420/7,449
Green Party	greenparty.ie	388(1999-2014)	517009	4,830,354/none
Anti Austerity Alliance	antiausterityalliance.ie	9(2014)	16317	3,199,124/none
Department of Agriculture	agriculture.gov.ie	580(2003-2015)	923749	37,329/187(gov.ie)
Teagasc	teagasc.ie	360(1998-2014)	1260273	303,114/3,330
Public Policy	publicpolicy.ie	22(2007-2015)	144127	1,175,738/8,030
NESC	nesc.ie	92(1999-2014)	16003	4,254,207/none

🟥 **Blog** ⬛ **Political Party** 🟦 **Government Agency** 🟦 **Policy Group**

Sample Websites

Name	Website	Wayback Machine Records	Total links	Rank(Global & Ireland)
Boards	boards.ie	1454(2000-2015)	35638779	5,947/21
Politics	politics.ie	624(2002-2015)	109415	73,104/361
The journal	thejournal.ie	1905(2010-2015)	None	10,317/54
Ask About Ireland: Environment	askaboutireland.ie/enfo/irelands-environment	32(2012-2014)	34159	216,744/14,165 (.ie/Australia)
Youth Development	youthdeved.ie	26(2007-2015)	76594	6,946,811/none
Science Spin	sciencespin.com	64(2003-2015)	14493	4,900,513/none
Trócaire	trocaire.org	621(1996-2015)	635640	773,948/10,349
Irish Aid	irishaid.ie	38(2009-2015)	33547	1,262,919/758,781(US)
Concern	concern.net	543(2001-2015)	774094	502,922/31,283(Pakistan)
Christian Aid	christianaid.ie	60(2004-2014)	36460	7,346,456/none

🟥 **Discussion and Magazine** 🟪 **Education** 🟩 **Charity**

WayBackMachine

- WBM is accessed by URL and by date of snapshot
- WBM provides PRESERVATION and ACCESSIBILITY but not so good on access, or promotion … no search facility (which is not unreasonable)

Archive.org

- WBM is enabled by modern technology (storage) which has also enabled other forms of CH preservation

Sports

- Sports / games are a form of intangible cultural heritage
- There are +3,000 traditional sports and games worldwide
- Rugby, Soccer, cricket, hockey are examples of the well-known, but …

Longe Paume

- A form of tennis with virtual nets

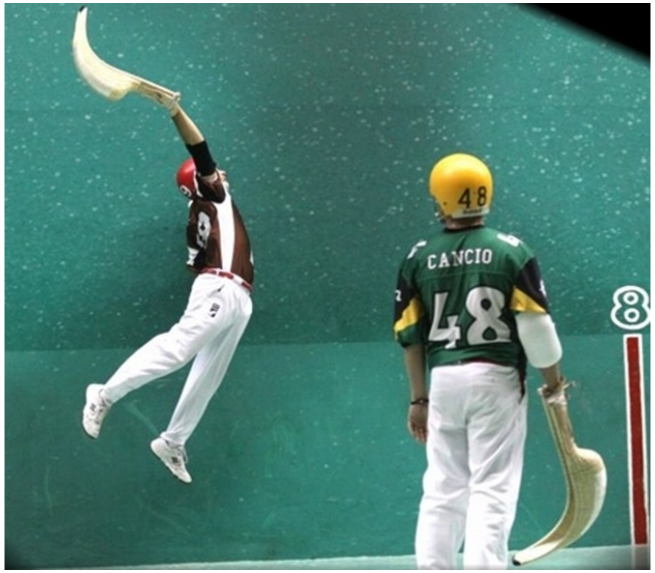
- Played in Northern France

Tamburello

- A form of vollyball but with no nets, and played with tambourines

- Played in Japan

Jai-Alai / Pelota



- The fastest game in the world
- Like handball but with a thresher and played as a team !
- Played in Basque

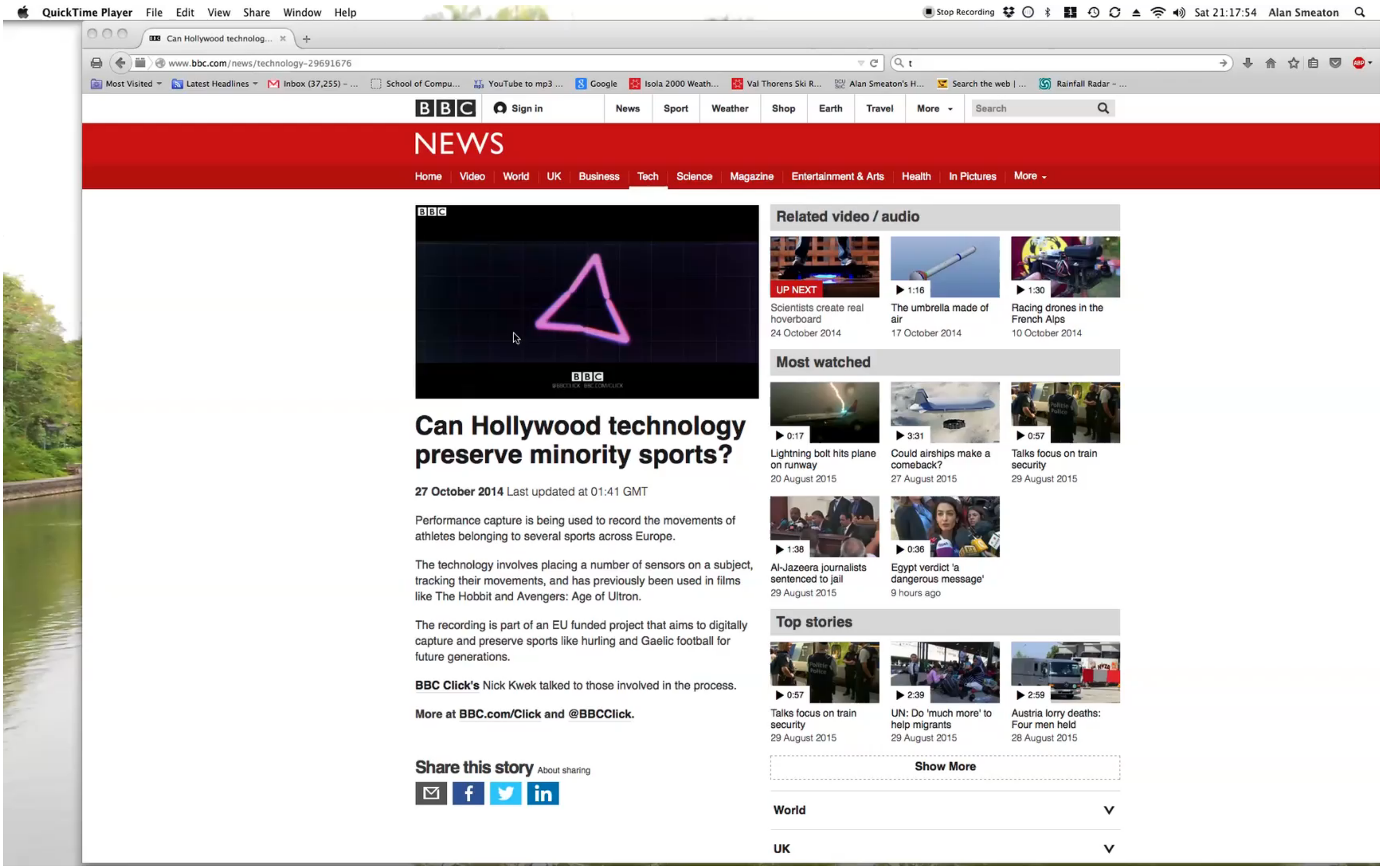
GAA Football and Hurling ;-)

- National game of Ireland

sportsfile 30 August 2015; Kevin McManamon, Dublin, is tackled by Lee Keegan, Mayo. GAA Football All-Ireland Senior Championship, Semi-Final, Dublin v Mayo, Croke Park, Dublin. Picture credit: Ramsey Cardy / SPORTSFILE

sportsfile 30 August 2015; Diarmuid Connolly, Dublin, tussles with Diarmuid O'Connor, Mayo. GAA Football All-Ireland Senior Championship, Semi-Final, Dublin v Mayo, Croke Park, Dublin. Picture credit: David Maher / SPORTSFILE

All examples of endangered sports !



Can Hollywood technology preserve minority sports?

27 October 2014 Last updated at 01:41 GMT

Performance capture is being used to record the movements of athletes belonging to several sports across Europe.

The technology involves placing a number of sensors on a subject, tracking their movements, and has previously been used in films like The Hobbit and Avengers: Age of Ultron.

The recording is part of an EU funded project that aims to digitally capture and preserve sports like hurling and Gaelic football for future generations.

BBC Click's Nick Kwek talked to those involved in the process.

More at BBC.com/Click and @BBCClick.

Share this story About sharing

Related video / audio

UP NEXT
Scientists create real hoverboard
24 October 2014

1:16
The umbrella made of air
17 October 2014

1:30
Racing drones in the French Alps
10 October 2014

Most watched

0:17
Lightning bolt hits plane on runway
20 August 2015

3:31
Could airships make a comeback?
27 August 2015

0:57
Talks focus on train security
29 August 2015

1:38
Al-Jazeera journalists sentenced to jail
29 August 2015

0:36
Egypt verdict 'a dangerous message'
9 hours ago

Top stories

0:57
Talks focus on train security
29 August 2015

2:39
UN: Do 'much more' to help migrants
29 August 2015

2:59
Austria lorry deaths: Four men held
28 August 2015

Show More

World

UK

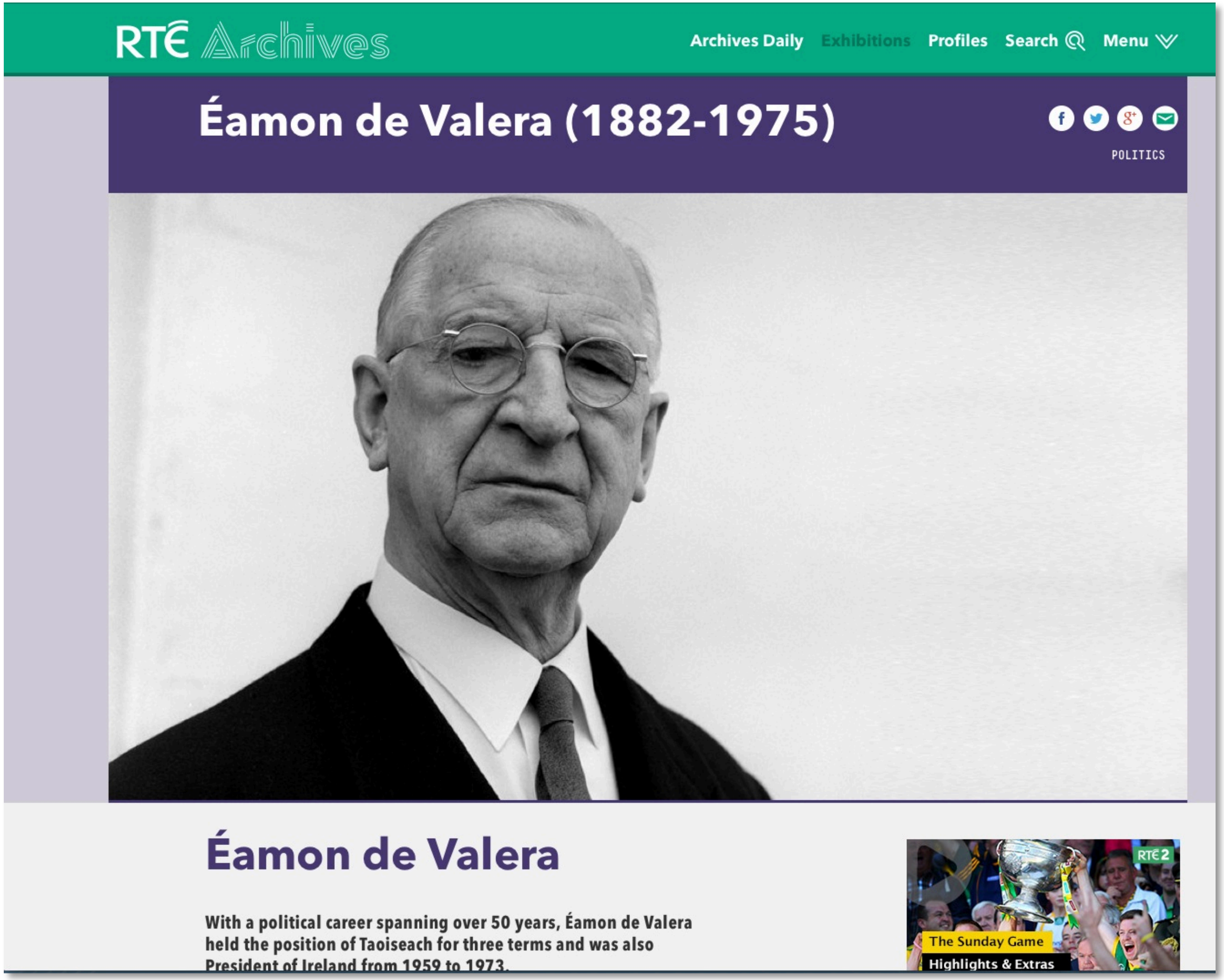
Promoting CH

- REPLAY addresses PRESERVATION, accessibility?, one form of ACCESS and a fair bit of PROMOTION by turning it into a game for public / children – outreach
- Promoting CH is difficult
- We like to think we like to engage in CH but its not easy to find interesting things, and we now like to find things
- Our attention span is reducing, especially younger generation, and we don't like guided tours
- Information consumption has changed, live TV is becoming rarer, YouTube is user control, we find news
- We want control over what we consume

Promoting CH

- At odds with this trend, digital CH is accessed in ways where others, not the consumer, are in control, just like TV and newspaper editors
- Example … TV archives

RTE.ie



Éamon de Valera (1882-1975)

Éamon de Valera

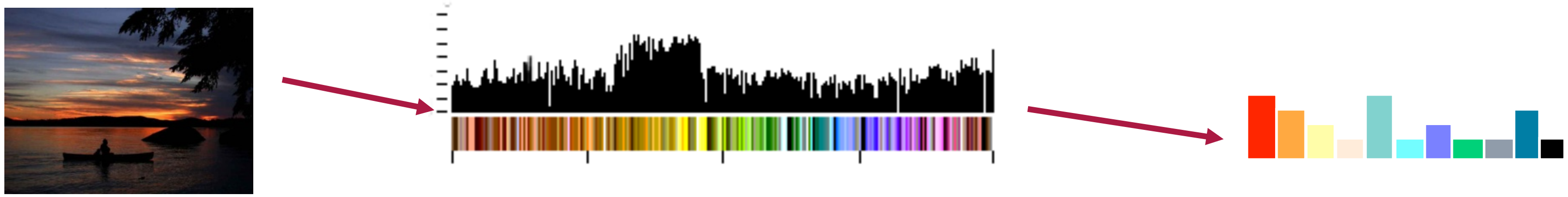
With a political career spanning over 50 years, Éamon de Valera held the position of Taoiseach for three terms and was also President of Ireland from 1959 to 1973.

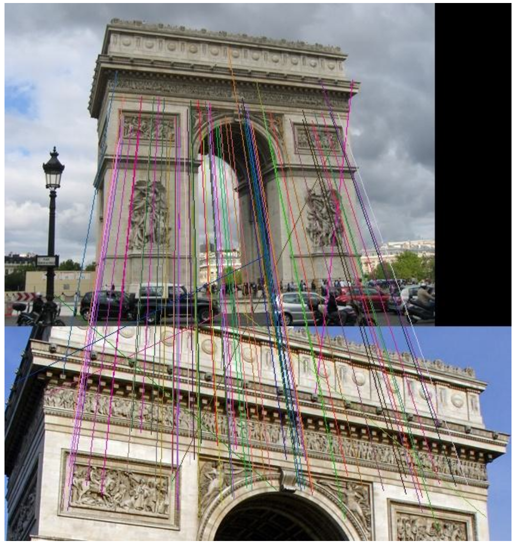
- **RTÉ Archive browsing is interesting, but could be enhanced …**
  - Locations could be built-in for visualisation, and navigation
  - Speech transcribed and searchable
  - Named entities (people, places, etc.) recognised and linked
  - … common characteristic is user control
- **User control for text navigation is easy … Google-like searching**
- **User control for visual navigation is not easy, we need to be able to automatically analyse imagery**

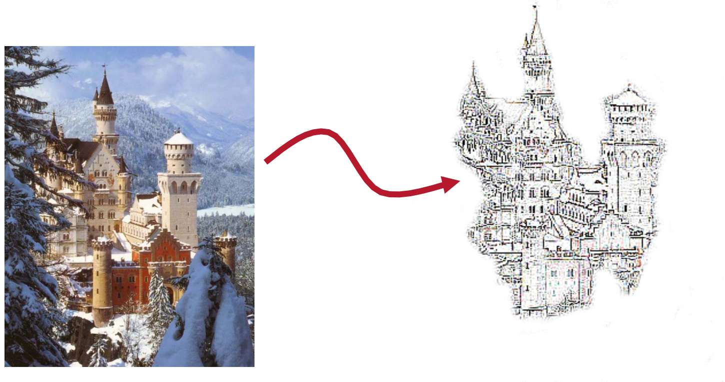
What does this mean?

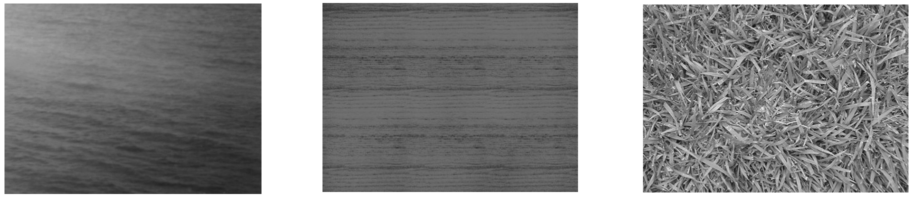
Or this?

Computer Vision / Image Processing



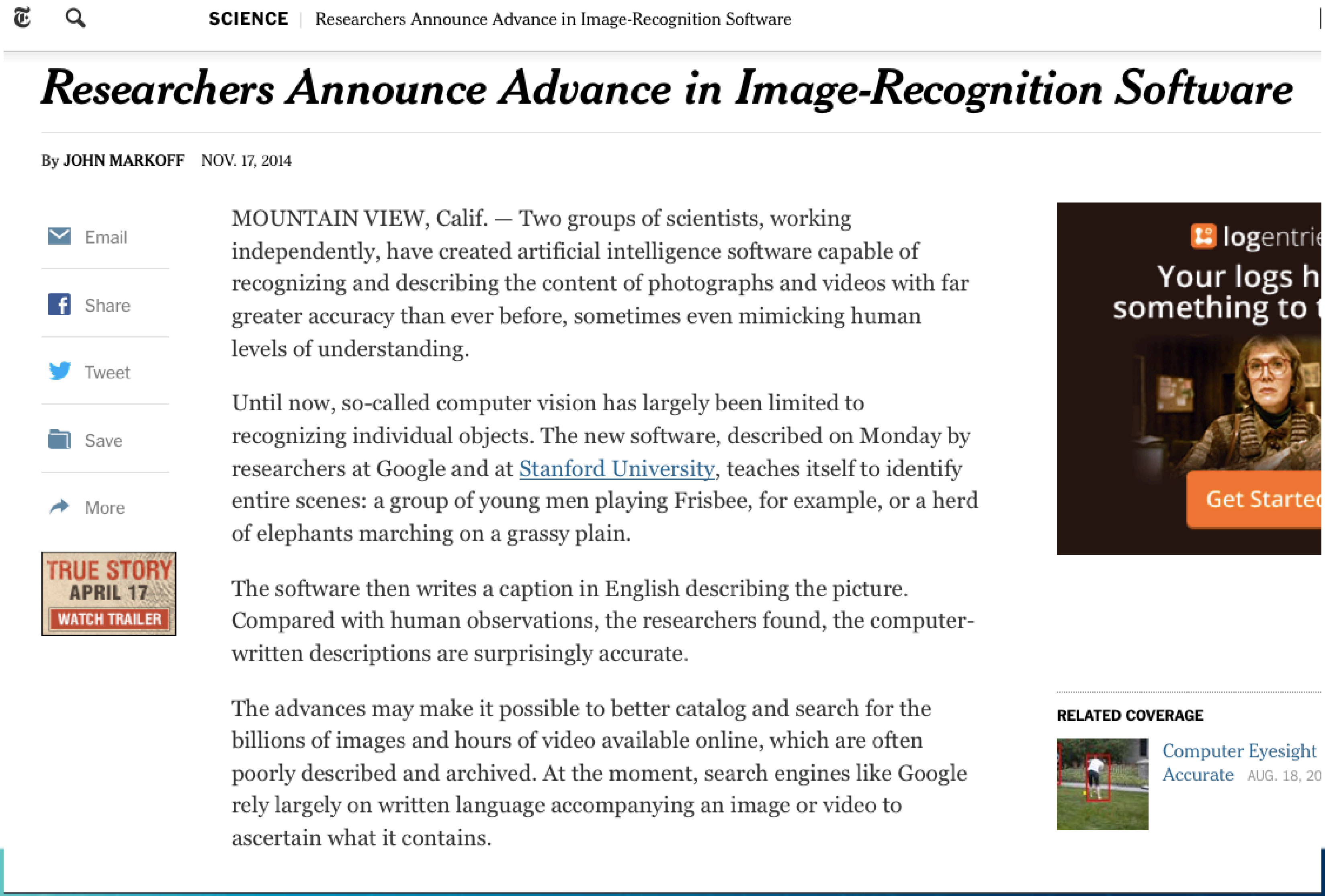






How good can a computer get at automatically annotating images without using any context at all ?

NYT, Nov 2014



SCIENCE | Researchers Announce Advance in Image-Recognition Software

Researchers Announce Advance in Image-Recognition Software

By JOHN MARKOFF NOV. 17, 2014

Email
Share
Tweet
Save
More

MOUNTAIN VIEW, Calif. — Two groups of scientists, working independently, have created artificial intelligence software capable of recognizing and describing the content of photographs and videos with far greater accuracy than ever before, sometimes even mimicking human levels of understanding.

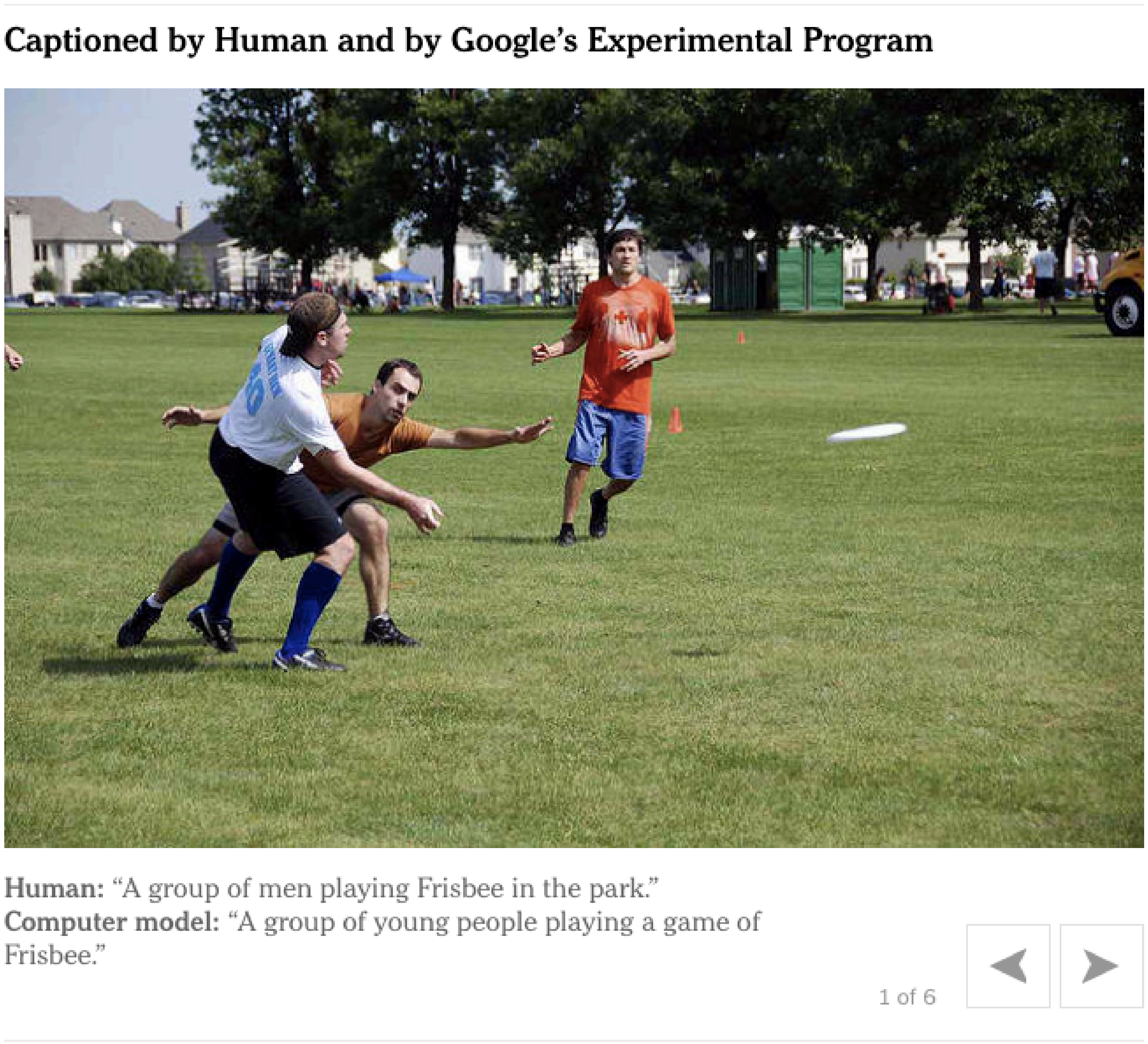
Until now, so-called computer vision has largely been limited to recognizing individual objects. The new software, described on Monday by researchers at Google and at Stanford University, teaches itself to identify entire scenes: a group of young men playing Frisbee, for example, or a herd of elephants marching on a grassy plain.

The software then writes a caption in English describing the picture. Compared with human observations, the researchers found, the computer-written descriptions are surprisingly accurate.

The advances may make it possible to better catalog and search for the billions of images and hours of video available online, which are often poorly described and archived. At the moment, search engines like Google rely largely on written language accompanying an image or video to ascertain what it contains.

RELATED COVERAGE

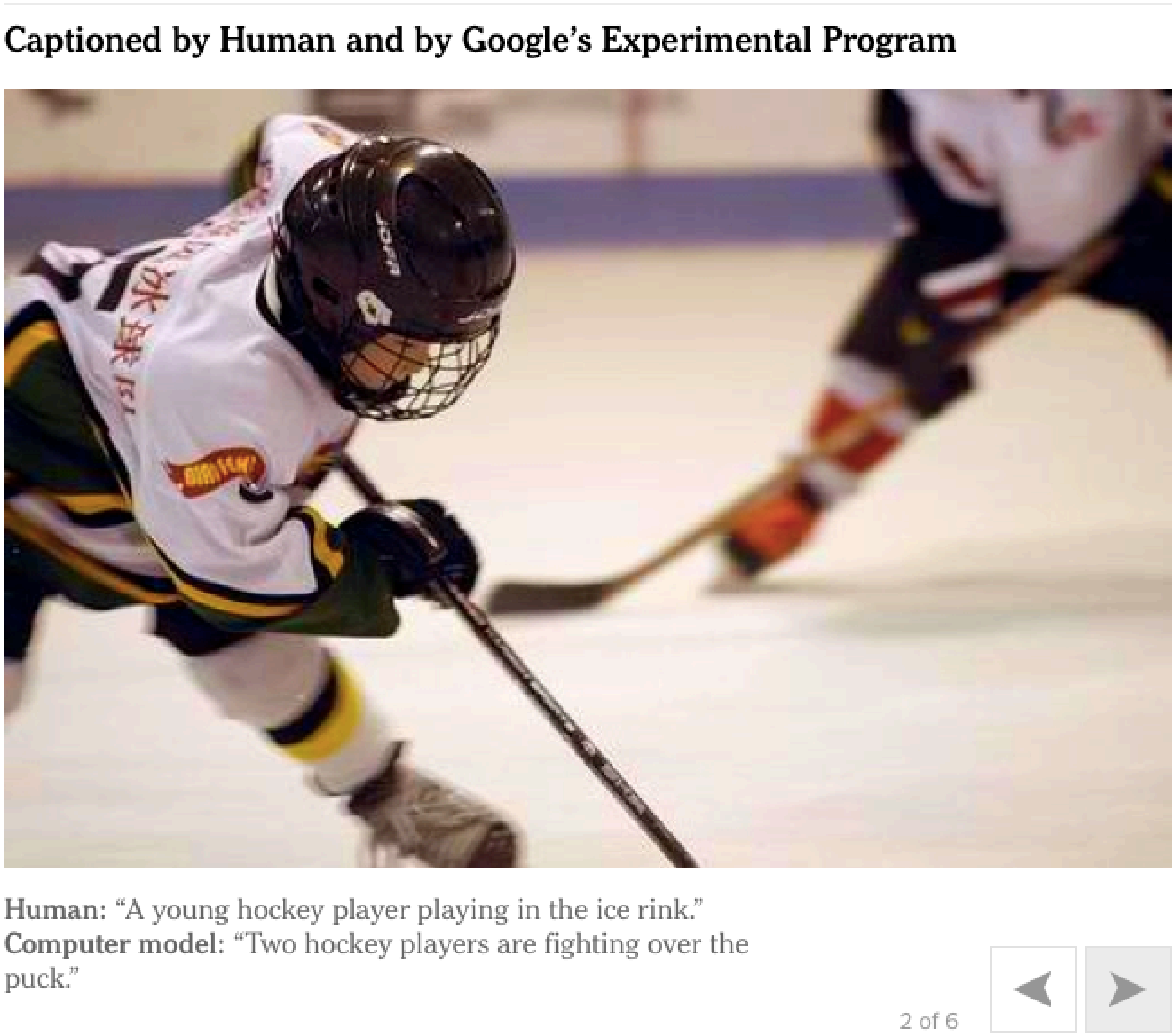
Computer Eyesight Accurate AUG. 18, 20



Captioned by Human and by Google’s Experimental Program

Human: “A group of men playing Frisbee in the park.”
Computer model: “A group of young people playing a game of Frisbee.”

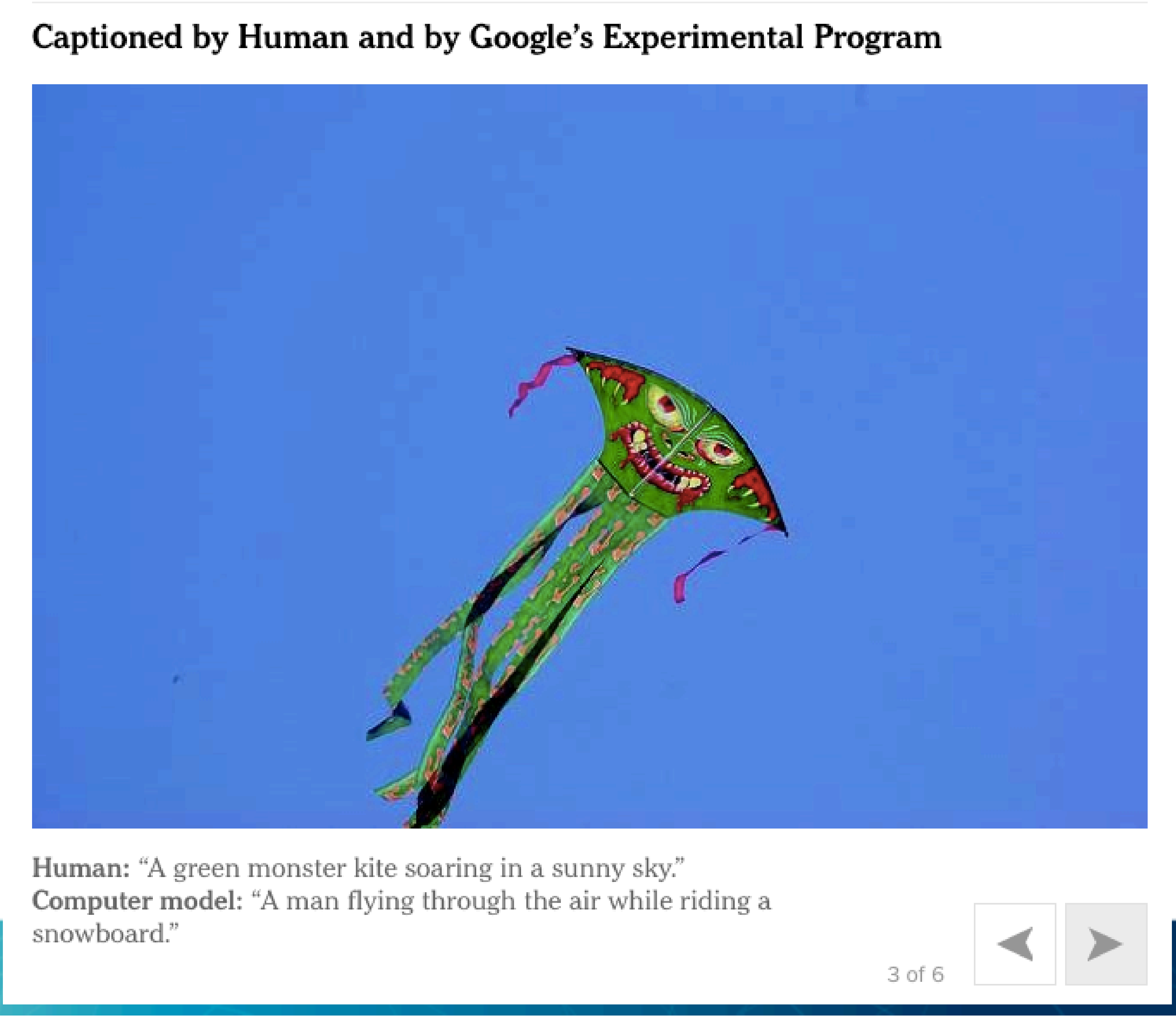
1 of 6



Captioned by Human and by Google's Experimental Program

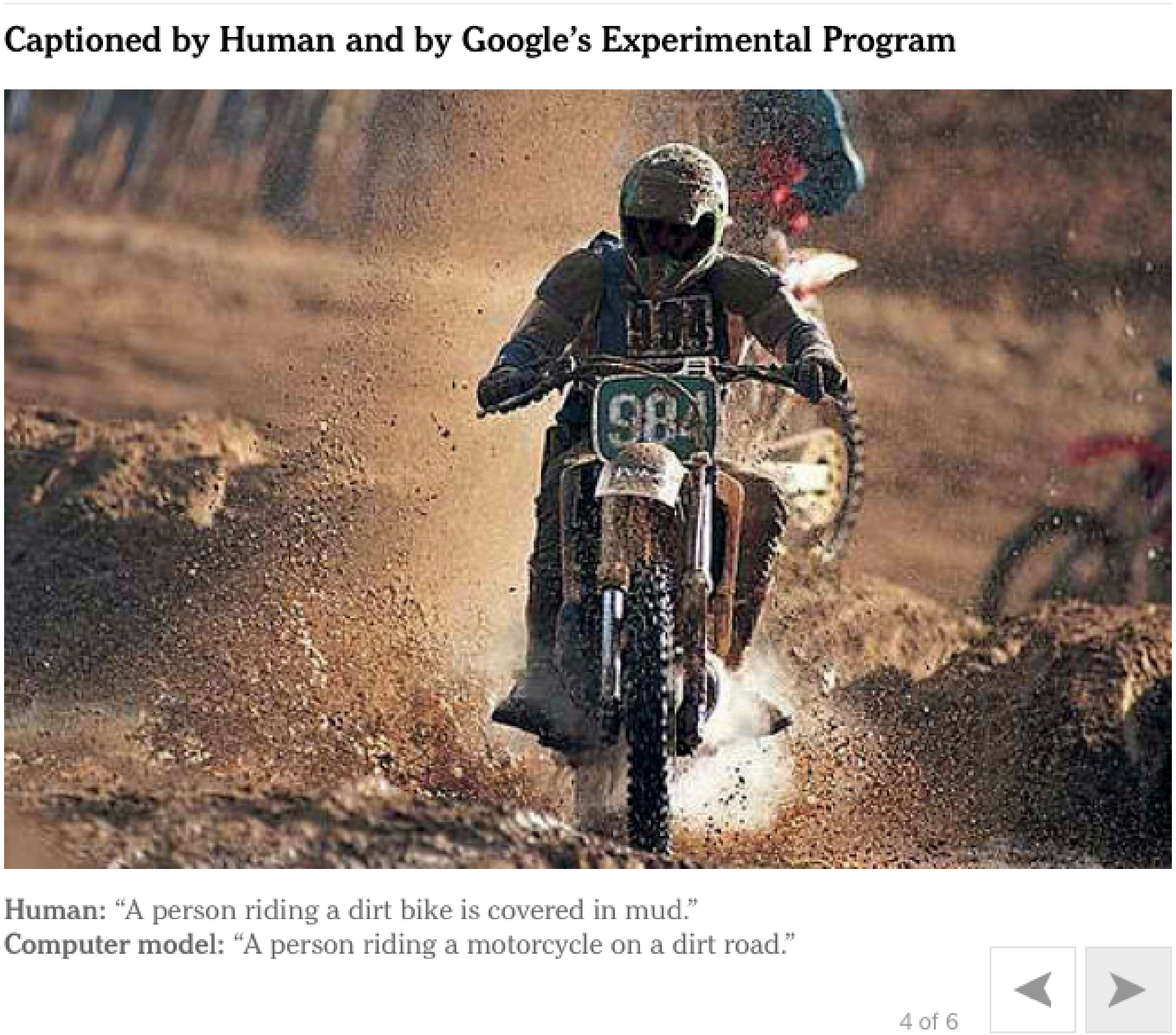
Human: "A young hockey player playing in the ice rink."
Computer model: "Two hockey players are fighting over the puck."

2 of 6



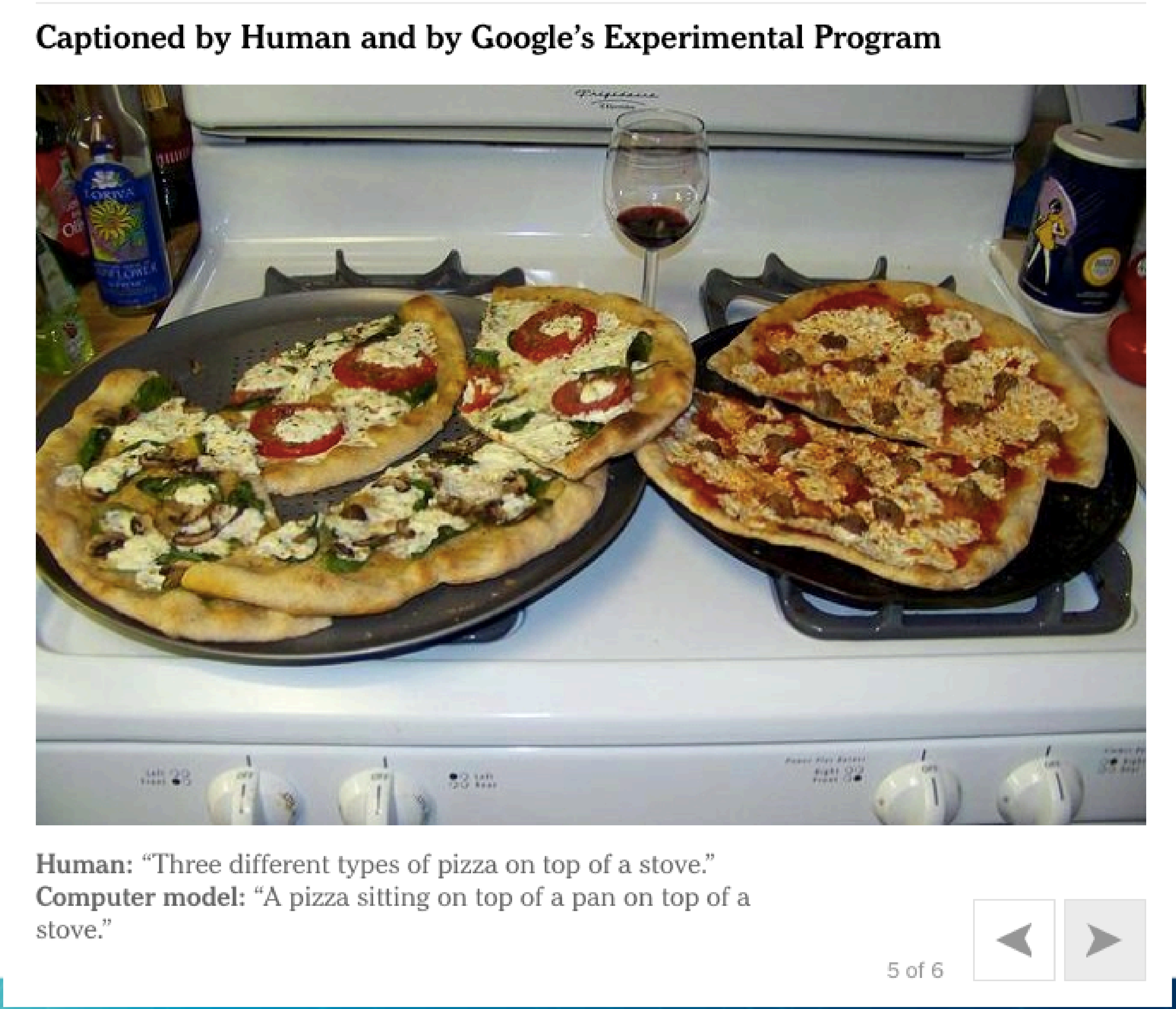
Captioned by Human and by Google's Experimental Program

Human: "A green monster kite soaring in a sunny sky."
Computer model: "A man flying through the air while riding a snowboard."



Captioned by Human and by Google's Experimental Program

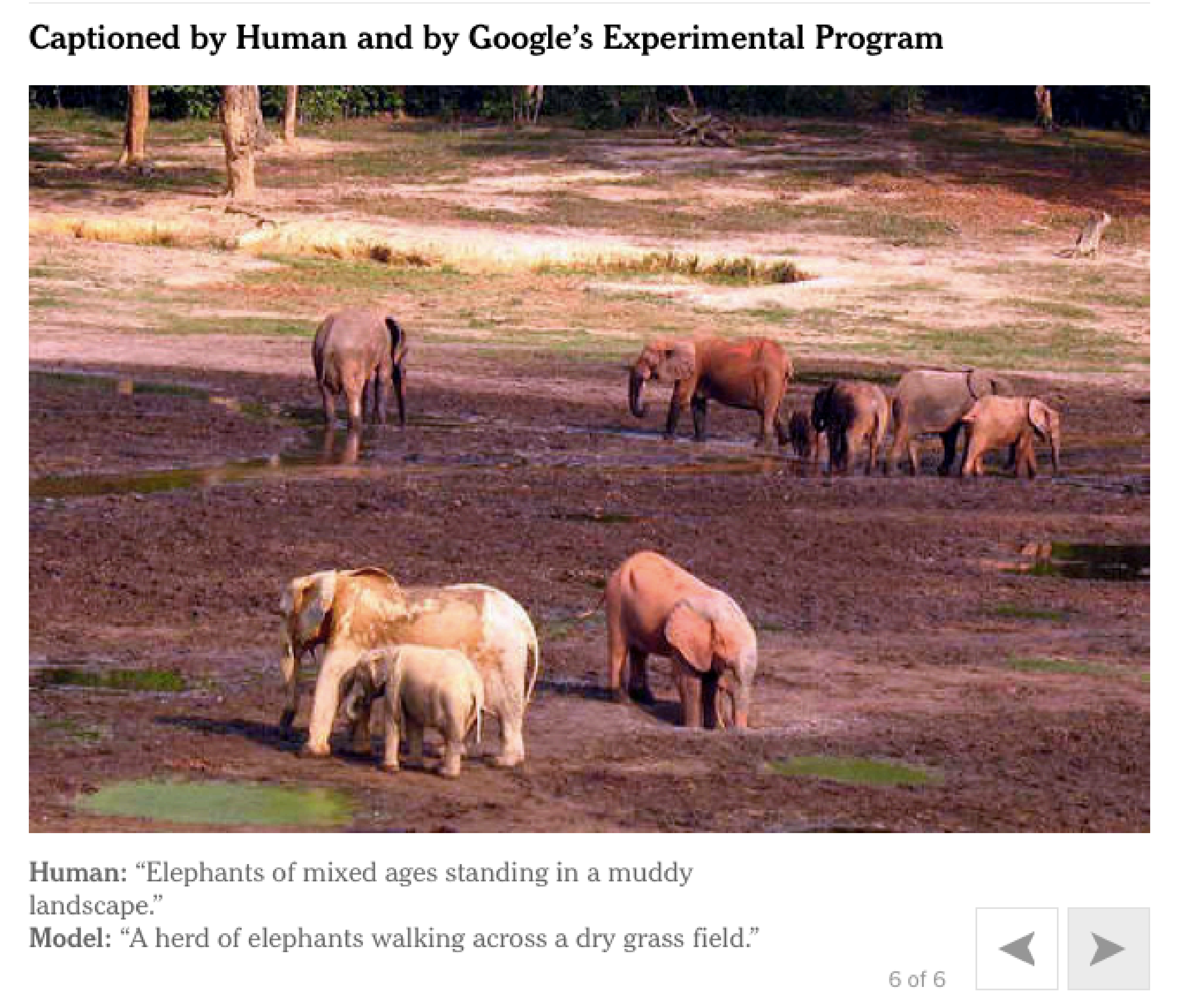
Human: "A person riding a dirt bike is covered in mud."
Computer model: "A person riding a motorcycle on a dirt road."



Captioned by Human and by Google’s Experimental Program

Human: “Three different types of pizza on top of a stove.”
Computer model: “A pizza sitting on top of a pan on top of a stove.”

5 of 6



Captioned by Human and by Google’s Experimental Program

Human: “Elephants of mixed ages standing in a muddy landscape.”
Model: “A herd of elephants walking across a dry grass field.”

How is captioning done ?

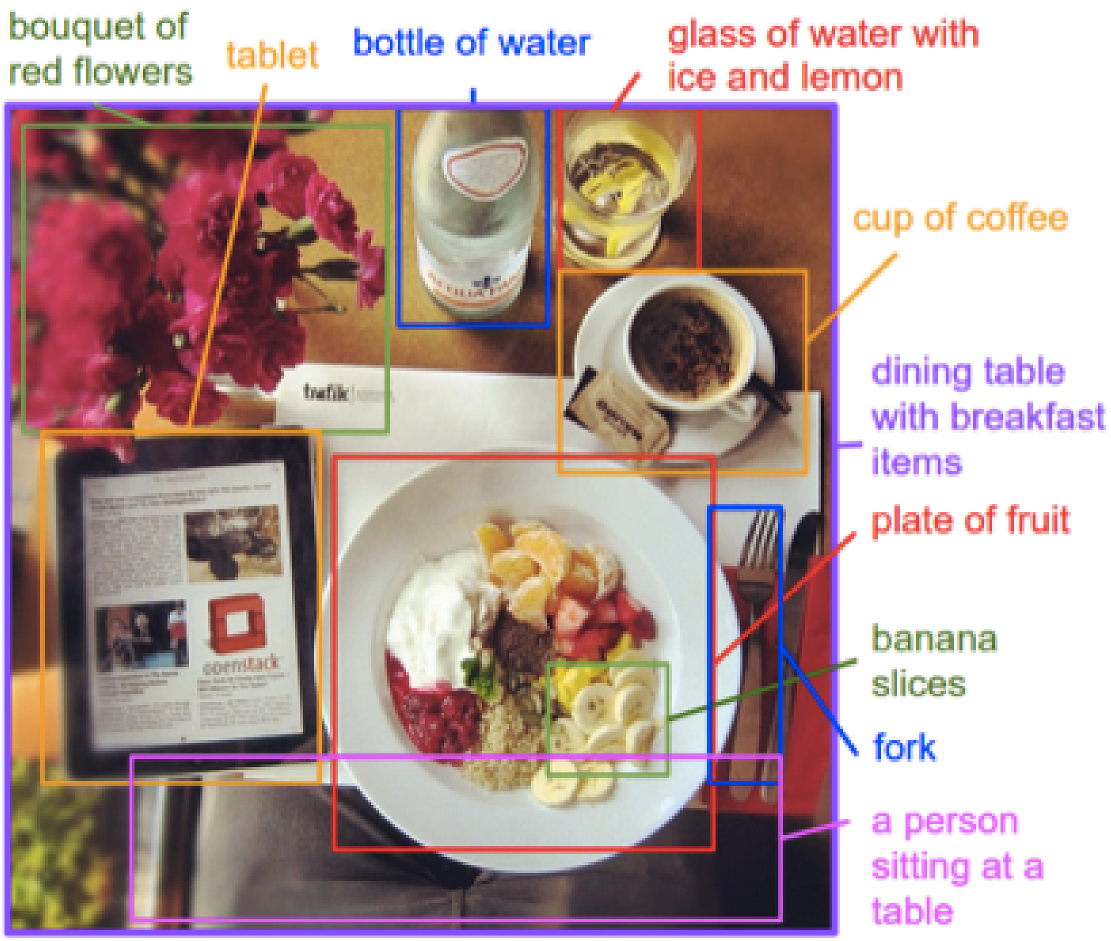

bouquet of red flowers
tablet
bottle of water
glass of water with ice and lemon
cup of coffee
dining table with breakfast items
plate of fruit
banana slices
fork
a person sitting at a table

Figure 1. Our model generates free-form natural language descriptions of image regions.

Deep Visual-Semantic Alignments for Generating Image Descriptions

Andrej Karpathy Li Fei-Fei
Department of Computer Science, Stanford University
{karpathy,feifeili}@cs.stanford.edu

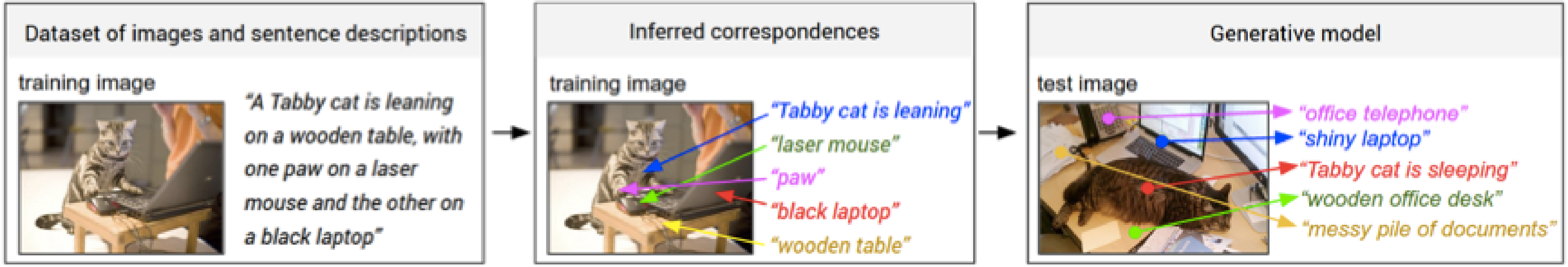


Figure 2. Overview of our approach. A dataset of images and their sentence descriptions is the input to our model (left). Our model first infers the correspondences (middle) and then learns to generate novel descriptions (right).

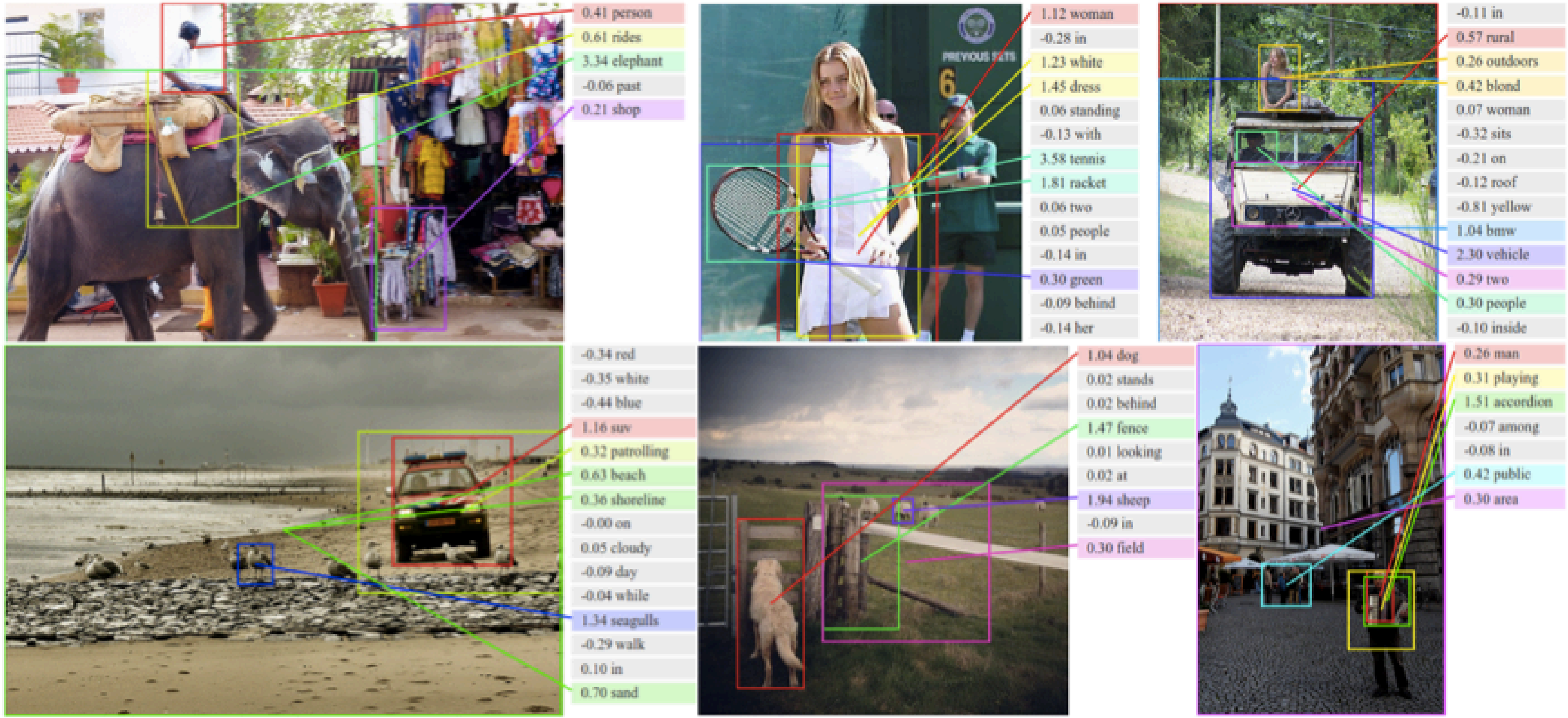
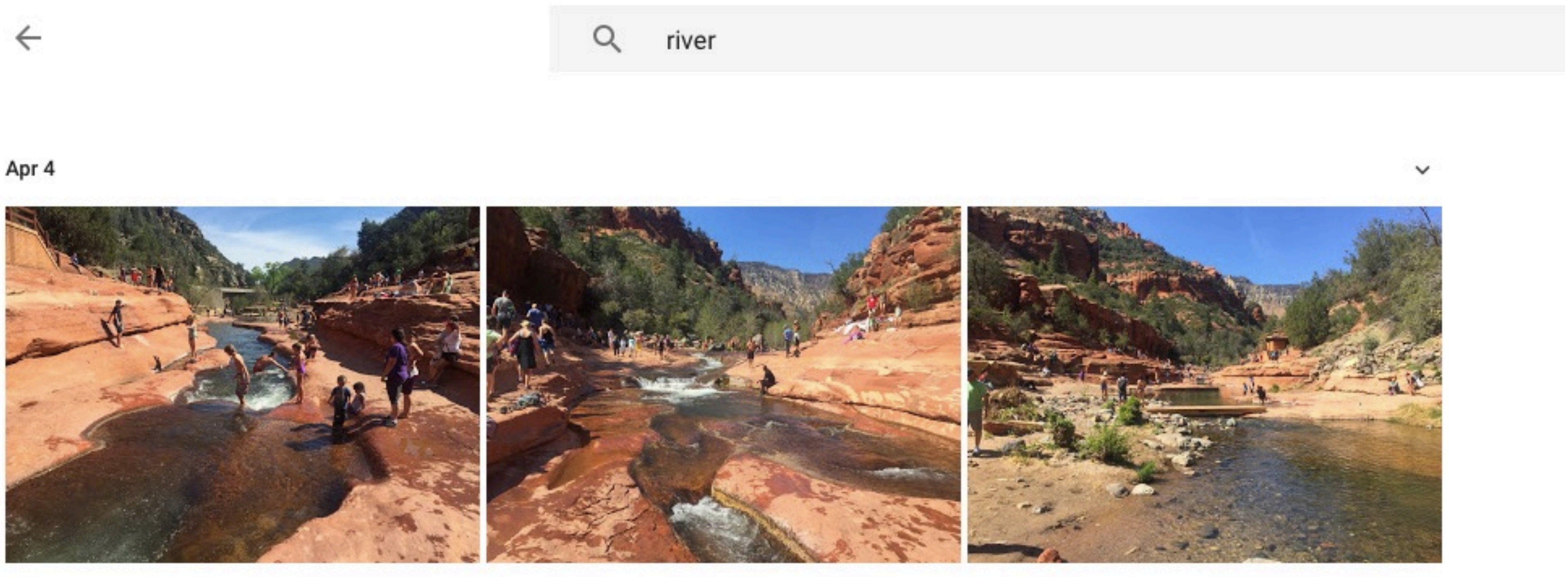
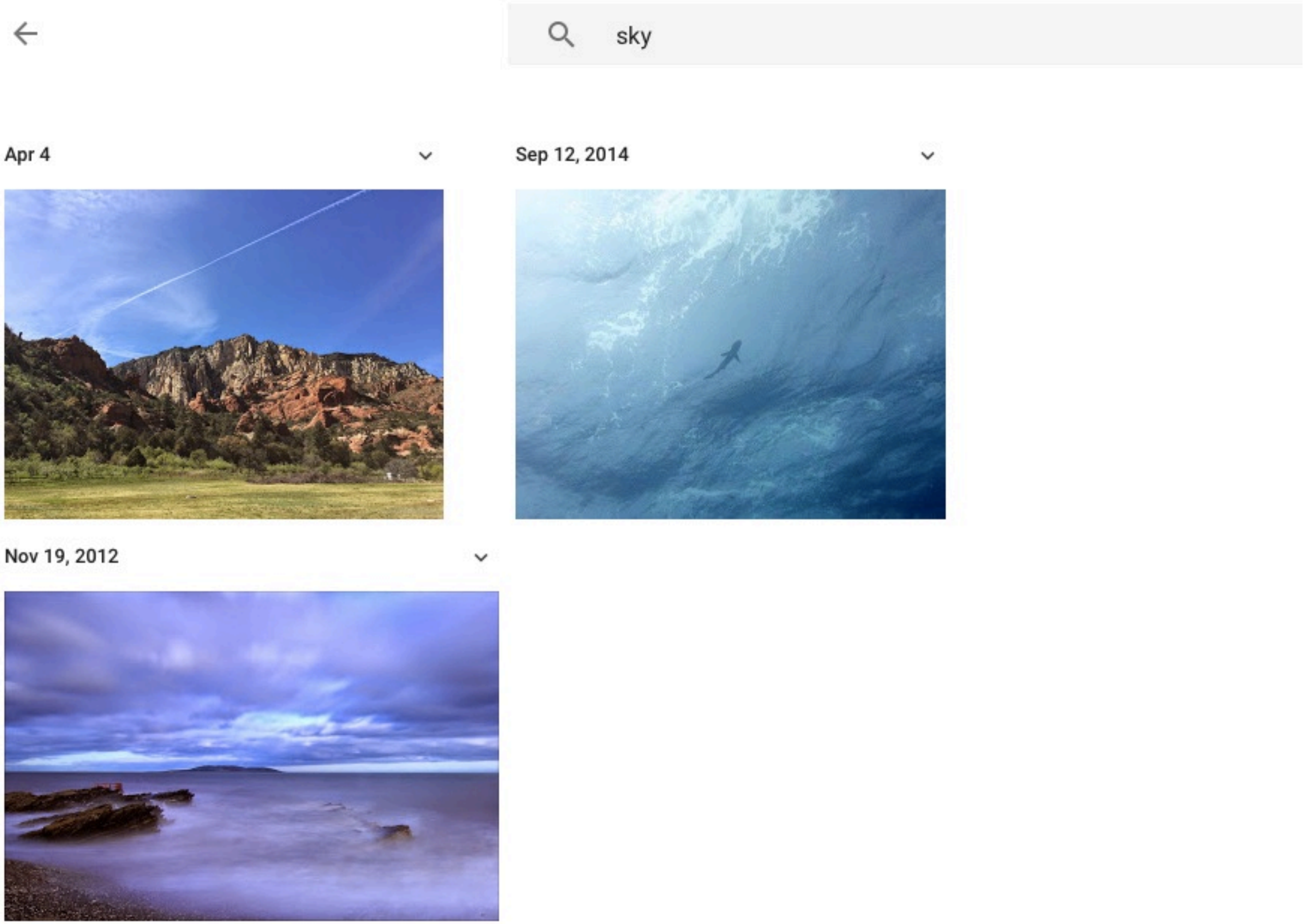


Figure 5. Example alignments predicted by our model. For every test image above, we retrieve the most compatible test sentence and visualize the highest-scoring region for each word (before MRF smoothing described in Section 3.1.4) and the associated scores ($v_i^T s_t$). We hide the alignments of low-scoring words to reduce clutter. We assign each region an arbitrary color.

Image Captioning ... Solved ?

- Computational processing needed for this is not scalable
- Domain knowledge data needed for this is not scalable
- Some major search engines are now introducing content based access to imagery … photos.google.com
- Based on pre-processing images for 00’s, 000’s, 0000’s of “concepts” … is that scalable ?


river
Apr 4

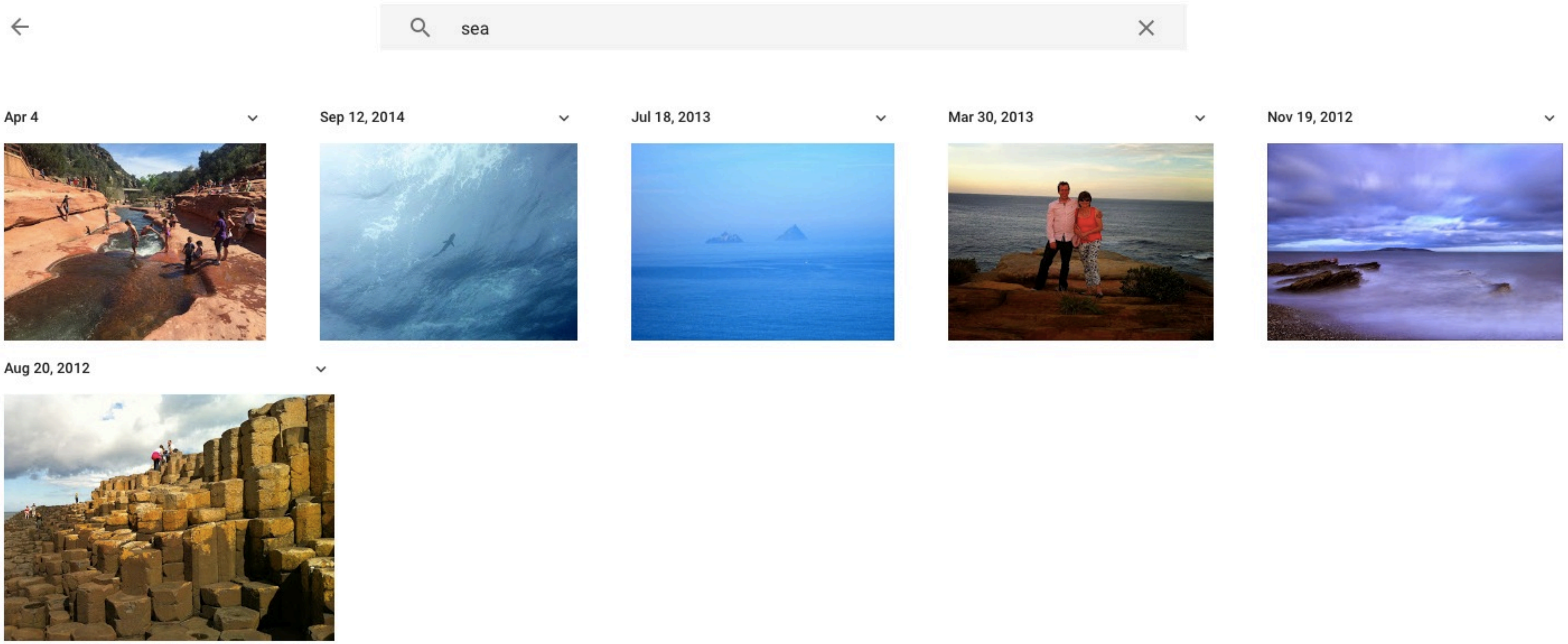


sky

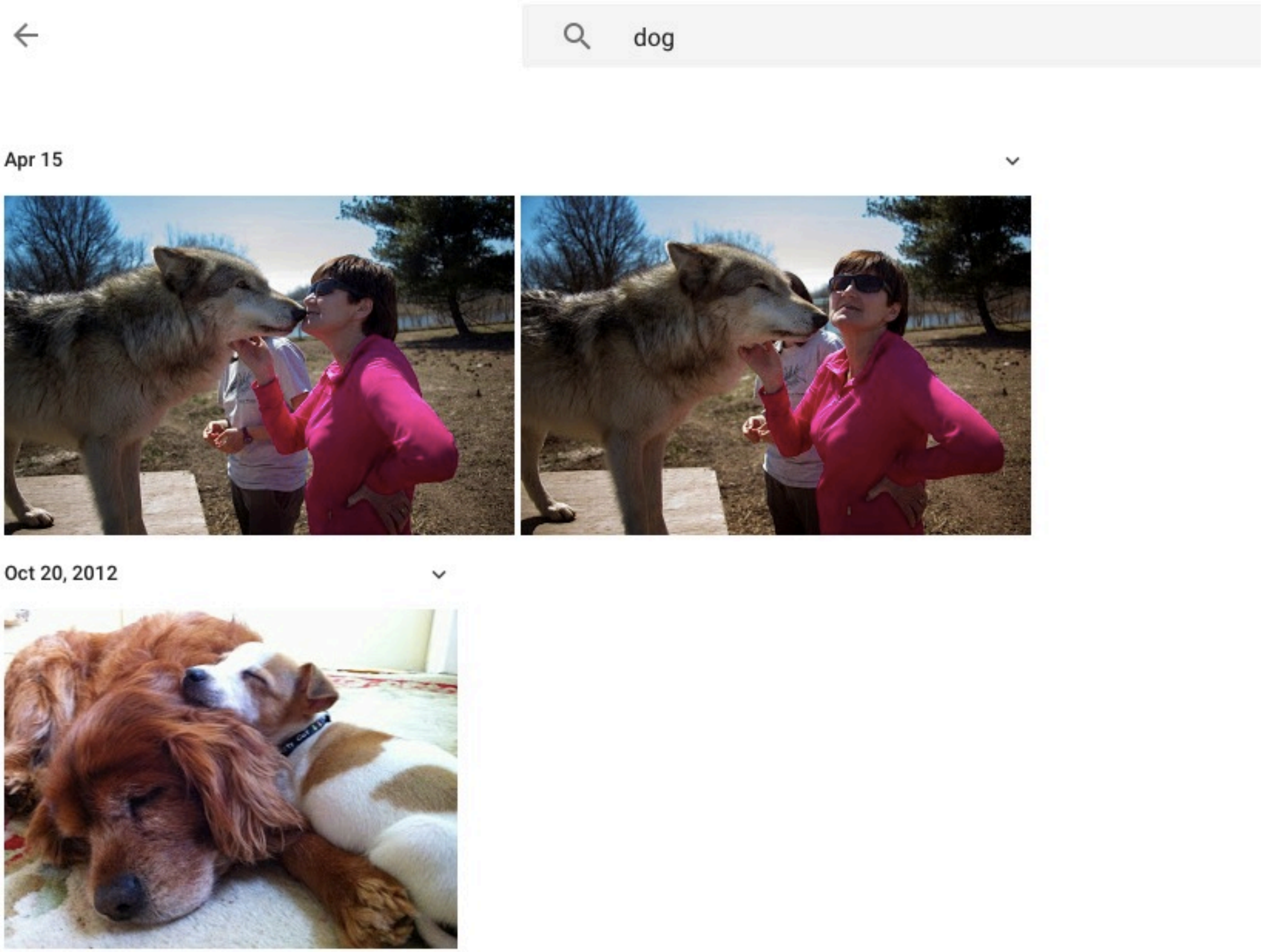
Apr 4

Sep 12, 2014

Nov 19, 2012



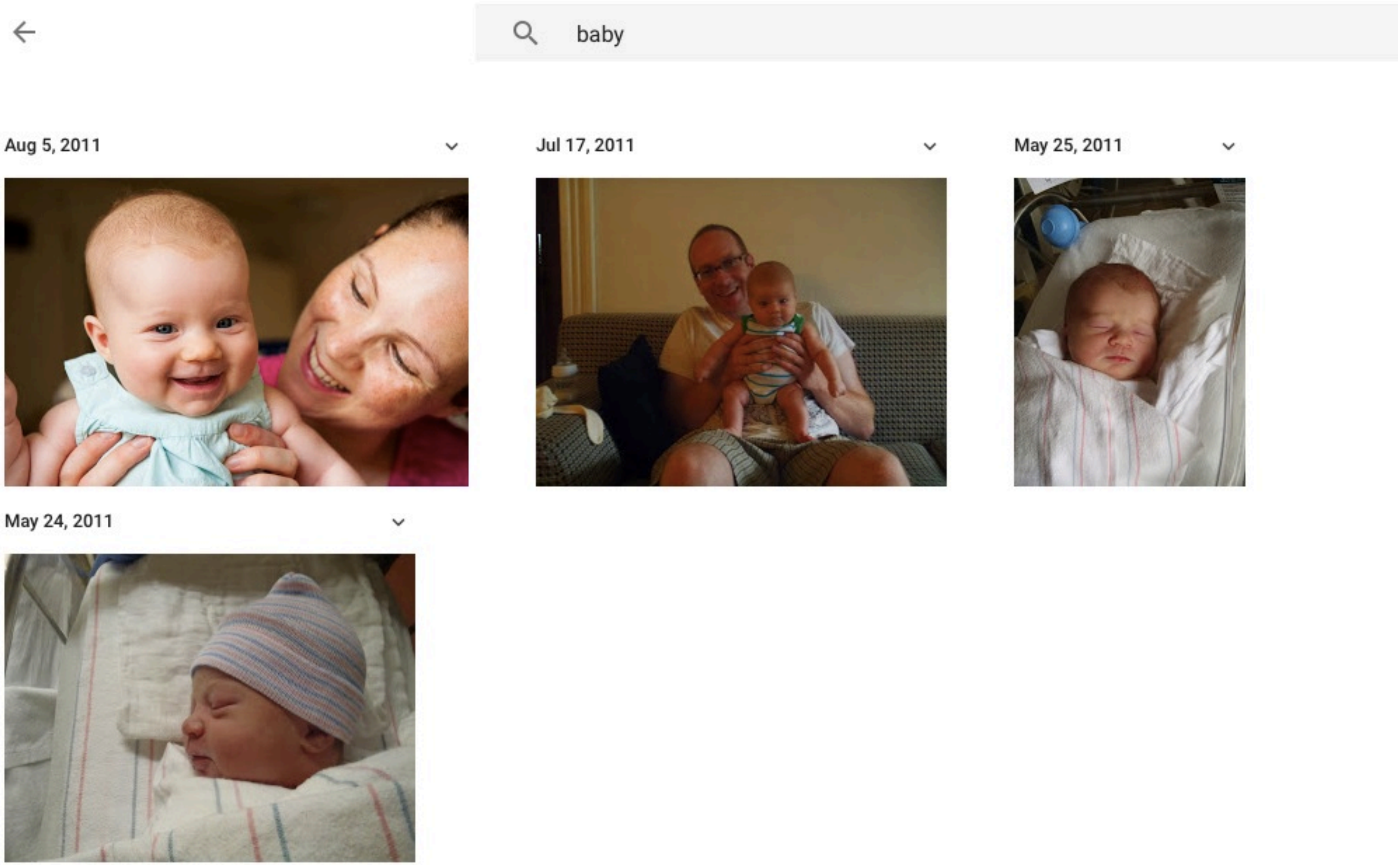
sea



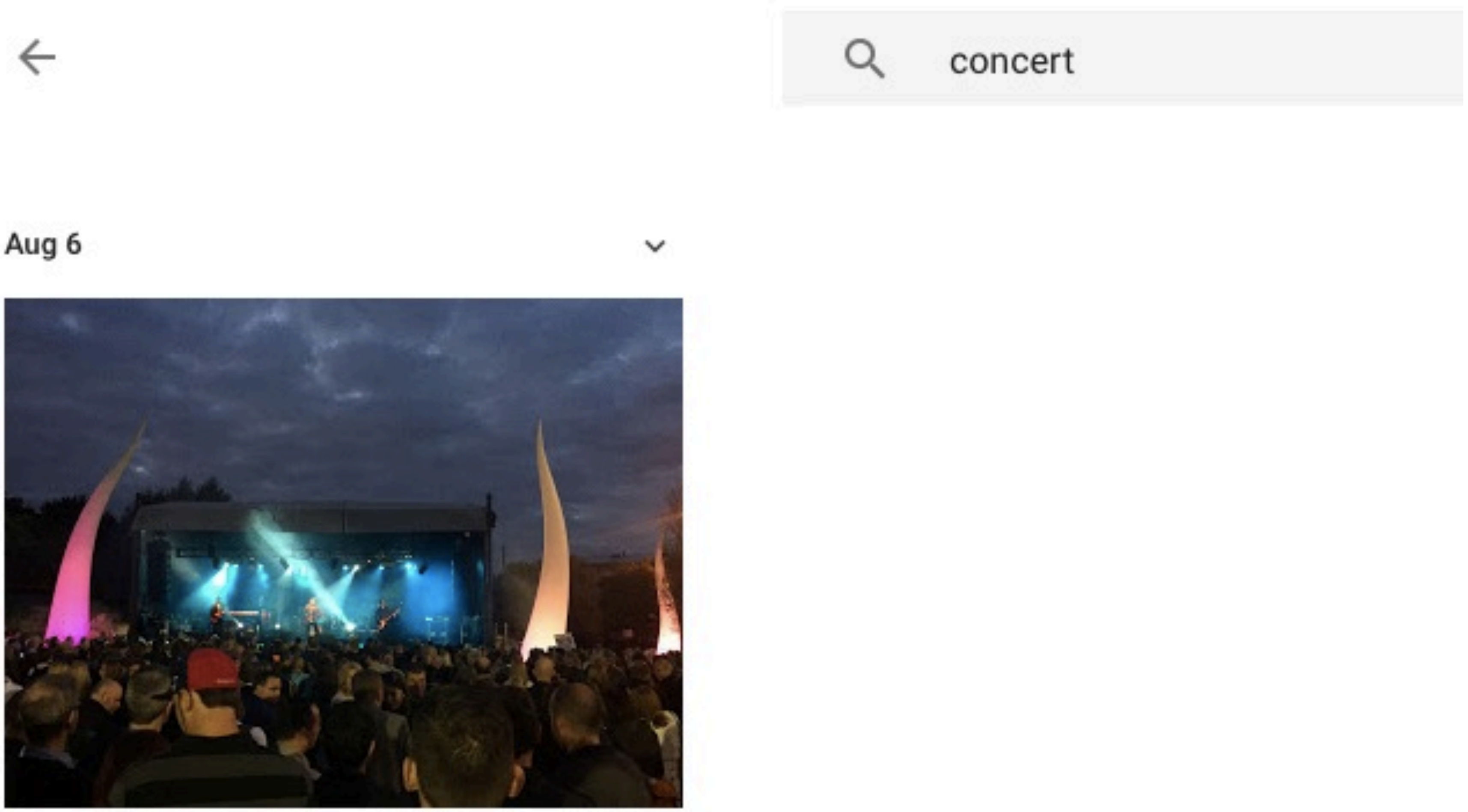
dog

Apr 15

Oct 20, 2012

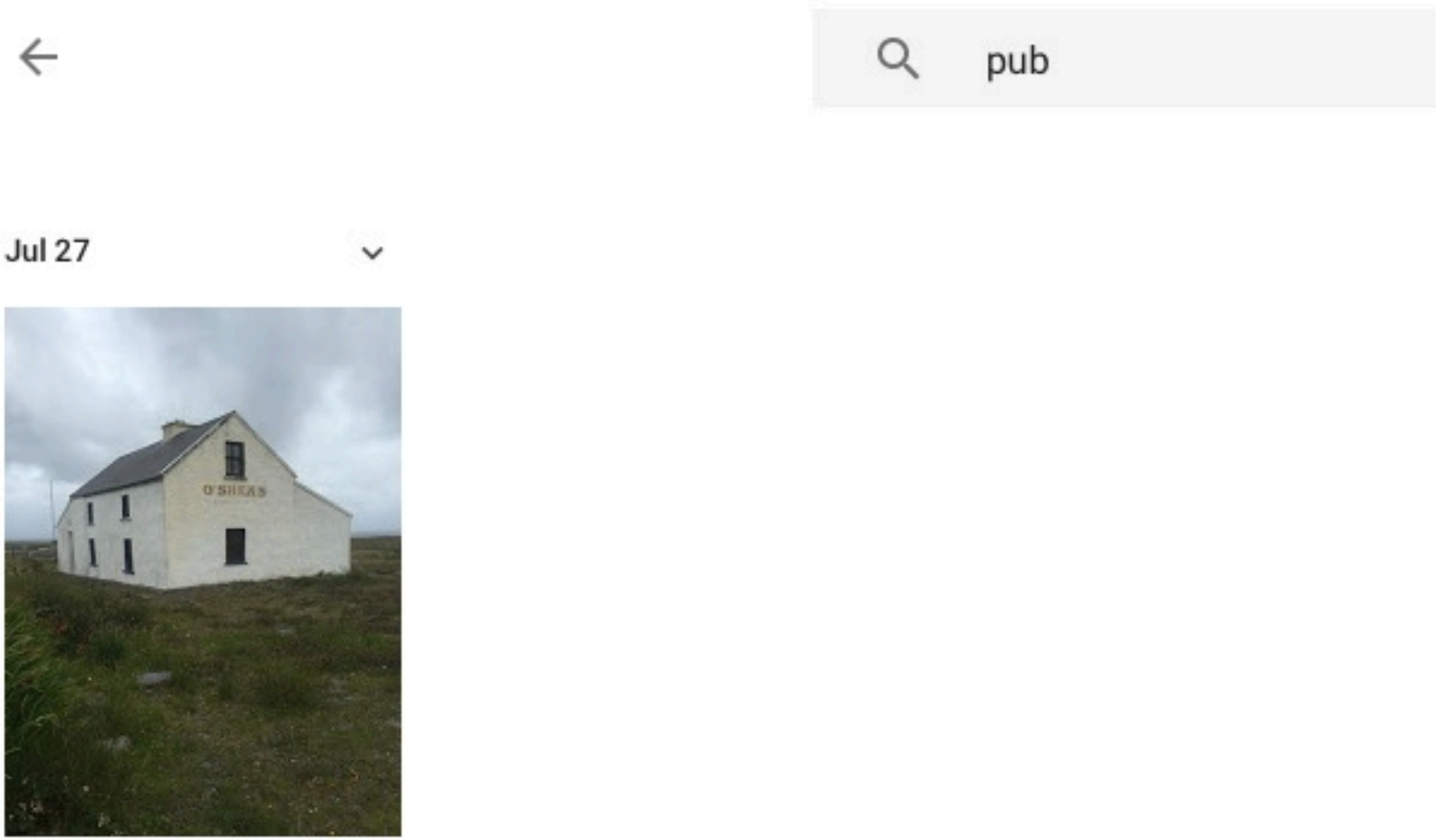


baby



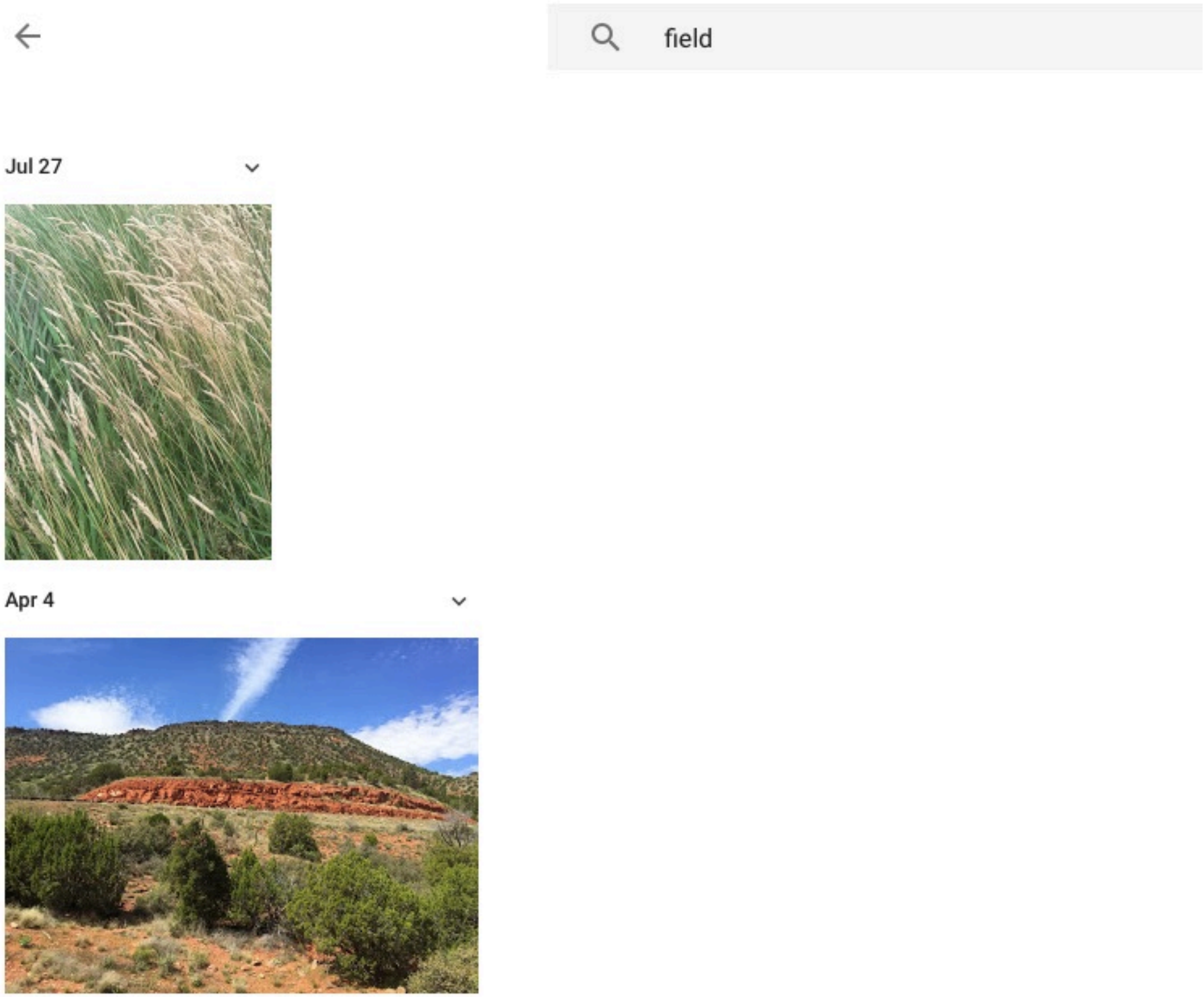
concert

Aug 6



pub

Jul 27



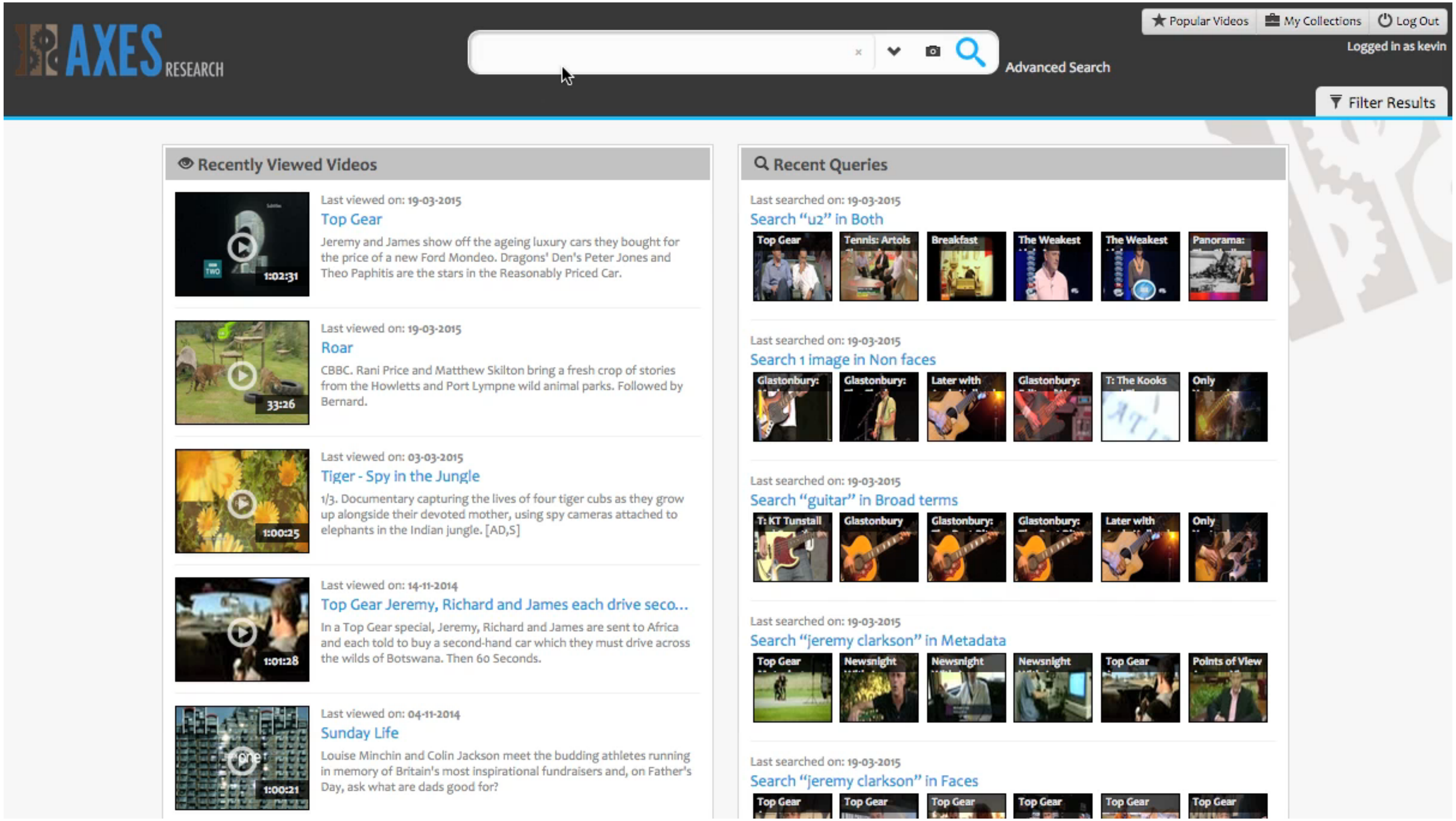
field

Jul 27

Apr 4

Searching visual libraries

- Scalable but only in a way
- These concept detectors are built by machine learning the differences between X and not X
- Requires much manual tagging and annotation
- An alternative to building these in advance is to built at query time … and where can we get lots of examples images of topic X … from Google of course !



Recently Viewed Videos

Last viewed on: 19-03-2015
Top Gear
Jeremy and James show off the ageing luxury cars they bought for the price of a new Ford Mondeo. Dragons' Den's Peter Jones and Theo Paphitis are the stars in the Reasonably Priced Car.

Last viewed on: 19-03-2015
Roar
CBBC. Rani Price and Matthew Skilton bring a fresh crop of stories from the Howletts and Port Lympne wild animal parks. Followed by Bernard.

Last viewed on: 03-03-2015
Tiger - Spy in the Jungle
1/3. Documentary capturing the lives of four tiger cubs as they grow up alongside their devoted mother, using spy cameras attached to elephants in the Indian jungle. [AD,S]

Last viewed on: 14-11-2014
Top Gear Jeremy, Richard and James each drive seco...
In a Top Gear special, Jeremy, Richard and James are sent to Africa and each told to buy a second-hand car which they must drive across the wilds of Botswana. Then 60 Seconds.

Last viewed on: 04-11-2014
Sunday Life
Louise Minchin and Colin Jackson meet the budding athletes running in memory of Britain's most inspirational fundraisers and, on Father's Day, ask what are dads good for?

Recent Queries

Last searched on: 19-03-2015
Search "u2" in Both

Last searched on: 19-03-2015
Search 1 image in Non faces

Last searched on: 19-03-2015
Search "guitar" in Broad terms

Last searched on: 19-03-2015
Search "jeremy clarkson" in Metadata

Last searched on: 19-03-2015
Search "jeremy clarkson" in Faces

Accessing Digital Cultural Heritage

- **So we are getting content-access … and there will be trickle-down into digital libraries for cultural heritage**
- **Digital Cultural Heritage in 10 years …**
  - New technologies, new media forms
  - New forms of Digital Cultural Heritage
  - New interaction modalities … tablets, augmented reality, brain sensing from wearables
  - New ways to navigate where we control
  - … but these are "just technology" changes and evolution
- **Biggest change will be in people, expectations, demands because People want to be in control**