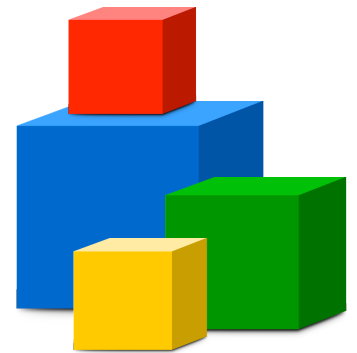


# Statistical learning from data as the ultimate agile development tool

Peter Norvig  
Google



# Agile Development Principles

- Customer satisfaction by rapid delivery
- Working software is **delivered frequently**
- Working software is the **measure of progress**
- Even late changes in requirements are welcomed
- Daily cooperation business people / developers
- Face-to-face conversation
- Projects built around motivated individuals
- Continuous attention to excellence of design
  - **Simplicity**
  - Self-organizing teams
  - Regular **adaptation** to changing circumstances

*see also: “Good Agile / Bad Agile,” Steve Yegge*

# Keep it Simple: More Data vs. Better Algorithms

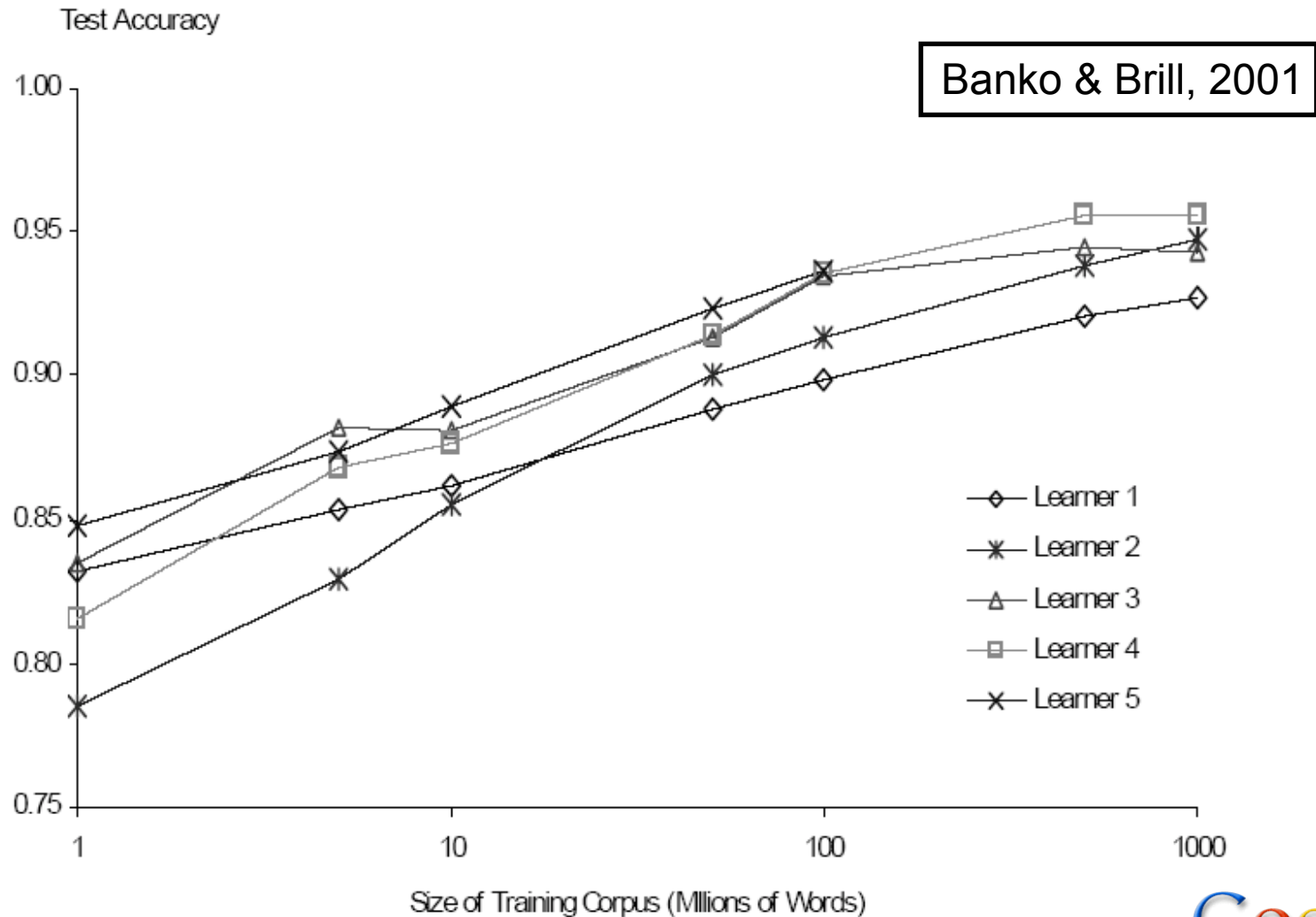


Figure 2. Learning Curves for Confusable Disambiguation



# Keep it Simple: More Data vs. Better Algorithms

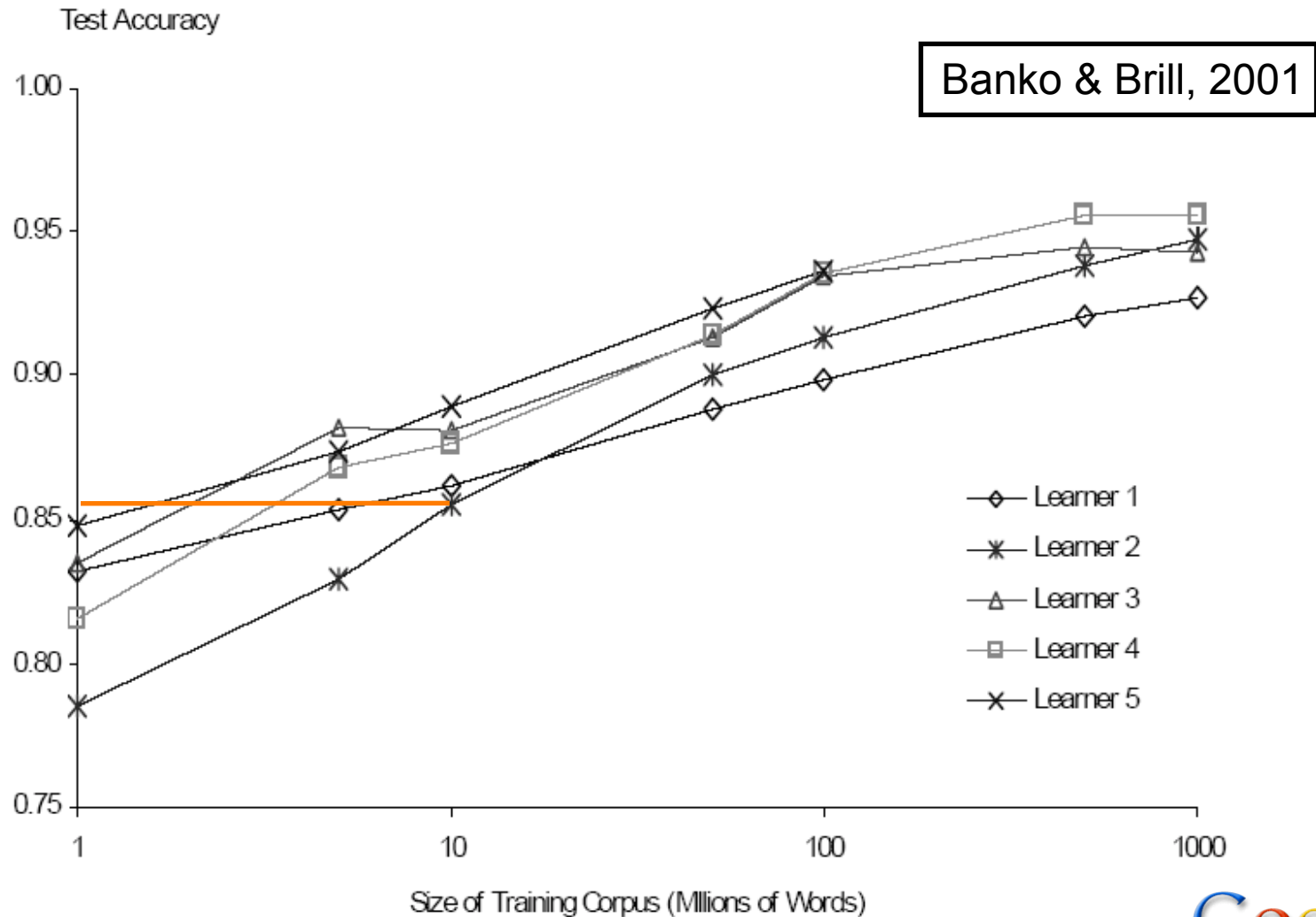


Figure 2. Learning Curves for Confusable Disambiguation



# Keep it Simple: More Data vs. Better Algorithms

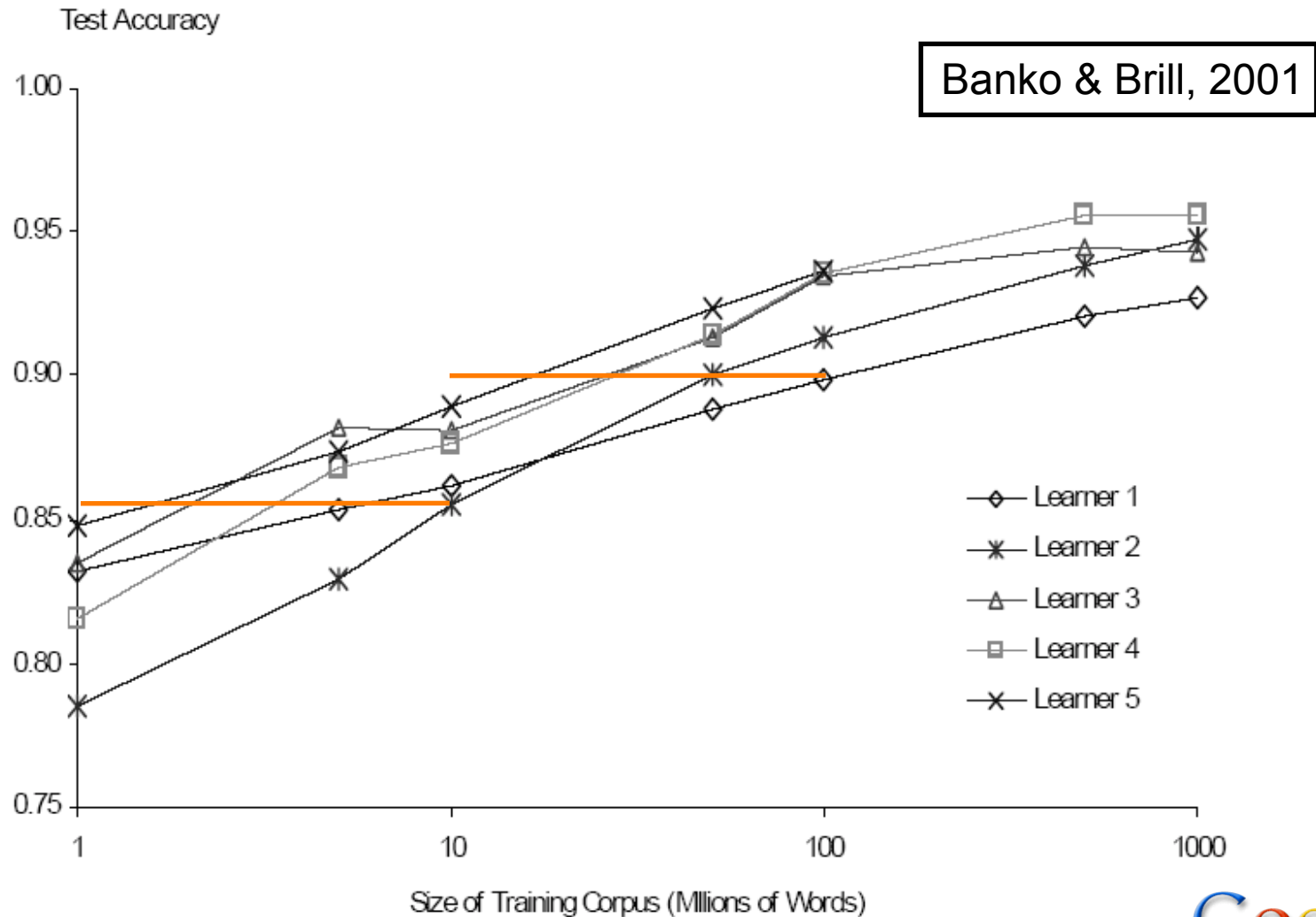


Figure 2. Learning Curves for Confusable Disambiguation



# Keep it Simple: More Data vs. Better Algorithms

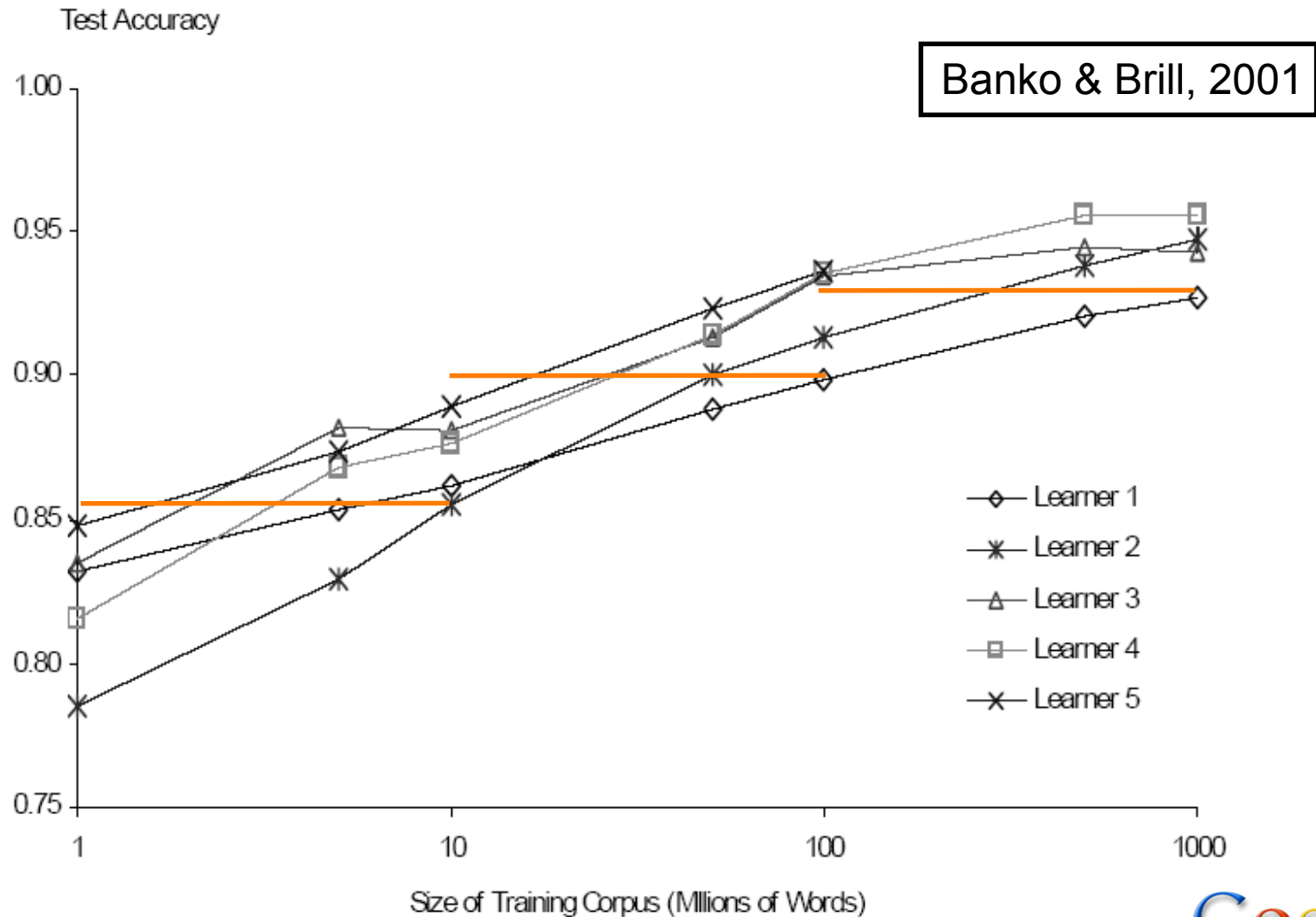


Figure 2. Learning Curves for Confusable Disambiguation



# Rational Programming

(or what to do when you don't know what to do)

Maximize Expected Utility:

$$action = \operatorname{argmax}_{a \in actions} EU(a)$$

$$EU(a) = \sum_{s \in Results(a)} P(s) \times U(s)$$

Learn/Approximate  $Results(a)$ ,  $P(s)$ ,  $U(s)$   
from big data sources

,





# Segmentation

**靳羽西中国新锐画家大奖**

# Segmentation

nowisthetimeforallgoodmentocometothe

# Segmentation

# Segmentation

Probability of a segmentation =  
 $P(\text{first word}) \times P(\text{rest})$

# Segmentation

Probability of a segmentation =  
 $P(\text{first word}) \times P(\text{rest})$

Best segmentation =  
one with highest probability

# Segmentation

Probability of a segmentation =  
 $P(\text{first word}) \times P(\text{rest})$

Best segmentation =  
one with highest probability

$P(\text{word}) =$   
estimated by counting

# Segmentation

segment("nowisthetime...")

$P_f("n") \times P_r("owisthetime...")$

$P_f("no") \times P_r("wisthetime...")$

$P_f("now") \times P_r("isthetime...")$

$P_f("nowi") \times P_r("sthetime...")$

...

# Segmentation

segment("nowisthetime...")

$f$	$P(f)$	$P(r)$	$= P(f) P(r)$
"n"	.003%	$\times 10^{-30}\%$	$= 10^{-34}\%$
"no"	.26%	$\times 10^{-26}\%$	$= 10^{-29}\%$
"now"	.23%	$\times 10^{-21}\%$	$= 10^{-24}\%$
"nowi"	$10^{-7}\%$	$\times 10^{-21}\%$	$= 10^{-30}\%$
...			



# Segmentation

```
from utils import Pw, product, memo
```

```
def splits(characters, longest=12):
```

```
    "All ways to split chars into a first word and remainder."
```

```
    return [(characters[:i], characters[i:])
```

```
            for i in range(1, 1+min(longest, len(characters)))]
```

```
def Pwords(words): return product(words, key=Pw)
```

```
@memo
```

```
def segment(text):
```

```
    "Best segmentation of text into words, by probability."
```

```
    return [ ] if (text=="") else (
```

```
        max([[first]+segment(rest) for first,rest in splits(text)],
```

```
            key=Pwords))
```

# Trained on 1.7B words English

# Trained on 1.7B words English

- 98% word accuracy. Typical errors:

# Trained on 1.7B words English

- 98% word accuracy. Typical errors:
- baseratesoughtto

# Trained on 1.7B words English

- 98% word accuracy. Typical errors:
- baseratesoughtto  
base rateesought to

# Trained on 1.7B words English

- 98% word accuracy. Typical errors:
- baseratesoughtto  
base ratee sought to
- smallandinsignificant

# Trained on 1.7B words English

- 98% word accuracy. Typical errors:
- baseratesoughtto  
base ratee sought to
- smallandinsignificant  
small and insignificant

# Trained on 1.7B words English

- 98% word accuracy. Typical errors:
- baseratesoughtto  
base ratee sought to
- smallandinsignificant  
small and in significant
- ginormousego



# Trained on 1.7B words English

- 98% word accuracy. Typical errors:
- baseratesoughtto  
base ratee sought to
- smallandinsignificant  
small and in significant
- ginormousego  
g in or mouse e go

# Mistakes were made

# Mistakes were made

whorepresents.com ⇒ [“who”, “represents”]

# Mistakes were made

whorepresents.com ⇒[“who”, “represents”]

therapistfinder.com ⇒[“therapist”, “finder”]

# Mistakes were made

whorepresents.com ⇒[“who”, “represents”]

therapistfinder.com ⇒[“therapist”, “finder”]

expertsexchange.com ⇒[“experts”, “exchange”]

# Mistakes were made

whorepresents.com ⇒ [“who”, “represents”]

therapistfinder.com ⇒ [“therapist”, “finder”]

expertsexchange.com ⇒ [“experts”, “exchange”]

penisland.com ⇒ error: expected [“pen”, “island”] g

# Mistakes were made

whorepresents.com ⇒ [“who”, “represents”]

therapistfinder.com ⇒ [“therapist”, “finder”]

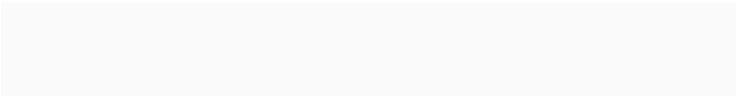
expertsexchange.com ⇒ [“experts”, “exchange”]

penisland.com ⇒ error: expected [“pen”, “island”] g

speedofart.net ⇒ [“speed”, “of”, “art”]

# Spelling

Mehran Sahami

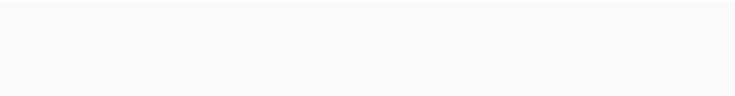




# Spelling

Typical word processor:

Mehran Sahami



# Spelling

Typical word processor:

Tehran Salami



- [..](#)
- [Accents.cc](#)
- [Accents.h](#)
- [Endings.cc](#)
- [Endings.h](#)
- [EndingsDB.cc](#)
- [Exact.cc](#)
- [Exact.h](#)
- [Fuzzy.cc](#)
- [Fuzzy.h](#)
- [Makefile.am](#)
- [Makefile.in](#)
- [Makefile.win32](#)
- [Metaphone.cc](#)**
- [Metaphone.h](#)
- [Prefix.cc](#)
- [Prefix.h](#)
- [Regexp.cc](#)
- [Regexp.h](#)
- [Soundex.cc](#)
- [Soundex.h](#)
- [Speling.cc](#)
- [Speling.h](#)
- [Substring.cc](#)
- [Substring.h](#)
- [SuffixEntry.cc](#)
- [SuffixEntry.h](#)
- [Synonym.cc](#)
- [Synonym.h](#)
- [htfuzzy.cc](#)

```
144
145     for (; *n && key.length() < MAXPHONEMELEN; n++)
146     {
147         /* Drop duplicates except for CC */
148         if (*(n - 1) == *n && *n != 'C')
149             continue;
150         /* Check for F J L M N R or first letter vowel */
151         if (same(*n) || *(n - 1) == '\0' && vowel(*n))
152             key << *n;
153         else
154         {
155             switch (*n)
156             {
157                 case 'B':
158                     /*
159                      * B unless in -MB
160                      */
161                     if (*(n + 1) || *(n - 1) != 'M')
162                         key << *n;
163                     break;
164                 case 'C':
165                     /*
166                      * X if in -CIA-, -CH- else S if in
167                      * -CI-, -CE-, -CY- else dropped if
168                      * in -SCI-, -SCE-, -SCY- else K
169                      */
170                     if (*(n - 1) != 'S' || !frontv(*(n + 1)))
171                     {
172                         if (*(n + 1) == 'I' && *(n + 2) == 'A')
173                             key << 'X';
174                         else if (frontv(*(n + 1)))
175                             key << 'S';
176                         else if (*(n + 1) == 'H')
177                             key << (((*(n - 1) == '\0' && !vowel
178                                     || *(n - 1) == 'S')
179                                     ? 'K' : 'X');
180                         else
181                             key << 'K';
182                     }

```

# Spelling with Statistical Learning

# Spelling with Statistical Learning

Probability of a spelling correction,  $c =$   
 $P(c \text{ as a word}) \times$   
 $P(\text{original is a typo for } c)$

# Spelling with Statistical Learning

Probability of a spelling correction,  $c =$   
 $P(c \text{ as a word}) \times$   
 $P(\text{original is a typo for } c)$

Best correction =  
one with highest probability

# Spelling with Statistical Learning

Probability of a spelling correction,  $c =$   
 $P(c \text{ as a word}) \times$   
 $P(\text{original is a typo for } c)$

Best correction =  
one with highest probability

$P(c \text{ as a word}) =$   
estimated by counting

# Spelling with Statistical Learning

Probability of a spelling correction,  $c =$   
 $P(c \text{ as a word}) \times$   
 $P(\text{original is a typo for } c)$

Best correction =  
one with highest probability

$P(c \text{ as a word}) =$   
estimated by counting

$P(\text{original is a typo for } c) =$   
proportional to number of changes



```
from utils import Pw, alphabet
```

```
def edits1(word):
```

```
    n = len(word)
```

```
    ## deletion, transposition, alteration, insertion
```

```
    return set([word[0:i]+word[i+1:] for i in range(n)] +  
               [word[0:i]+word[i+1]+word[i]+word[i+2:] for i in range(n-1)] +  
               [word[0:i]+c+word[i+1:] for i in range(n) for c in alphabet] +  
               [word[0:i]+c+word[i:] for i in range(n+1) for c in alphabet])
```

```
def known_edits2(word):
```

```
    return set(e2 for e1 in edits1(word) for e2 in edits1(e1) if e2 in Pw)
```

```
def known(words):
```

```
    return set(w for w in words if w in Pw)
```

```
def correct(word):
```

```
    candidates = (known([word]) or known(edits1(word)) or  
                 known_edits2(word) or [word])
```

```
    return max(candidates, key=Pw)
```

# Google™ Sets

Automatically create sets of items from a few examples.

Enter a few items from a set of things. ([example](#))

Next, press *Large Set* or *Small Set* and we'll try to predict other items in the set.

- ◆
- ◆
- ◆
- ◆
- ◆

[\(clear all\)](#)

Large Set

Small Set (15 items or fewer)

# Google Sets

Given “henri matisse,” “pablo picasso” find:

henri matisse, pablo picasso, vincent van gogh, claude monet, pablo picasso, salvador dali, edgar degas, paul cezanne, andy warhol, pierre auguste renoir, marc chagall, paul gauguin, mary cassatt, michelangelo, jackson pollock, camille pissarro, paul klee, georgia o'keeffe, norman rockwell, leonardo da vinci, edward hopper, Frida Kahlo, Rembrandt, Wassily Kandinsky, Rene Magritte, Francisco Goya, Raphael, Georges Seurat, Giotto, Henri Rousseau, Diego Rivera, Van Gogh, Botticelli, Johannes Vermeer, Ansel Adams, Monet, Titian

# Google Sets

Given “lion, tiger, bear” find:

bear, tiger, lion, elephant, monkey, giraffe, dog, cat, snake, horse, zebra, rabbit, wolf, dolphin, dragon, pig, frog, duck, cheetah, bird, cow, *cotton*, hippo, turtle, penguin, rat, gorilla, leopard, sheep, mouse, puppy, ox, rooster, fish, lamb, panda, *wood*, *musical*, *toddler*, fox, goat, deer, squirrel, koala, crocodile, hamster

# Google Sets

# Google Sets

Given “cat, man” find:

# Google Sets

Given “cat, man” find:

cat, man, ls, cp, rm, mkdir, mv, cd,  
pwd, rmdir, chmod, ln, grep, touch,  
find, ps, chown, df, less, tar, chgrp, du,  
sort, date, echo, kill, tail, wc, mount,  
sed, passwd, dd, head, vi, which, gzip,  
lpr, who, file, su, umount, diff, cut,  
uname, exit, basename, clear, sleep,  
whoami, mknod





# Google Sets

# Google Sets

**Tin Woodsman:** Some, but mostly lions and tigers and bears.

**Dorothy:** Lions?

**Scarecrow:** And tigers?

**Tin Woodsman:** And bears.

# Google Sets

**Tin Woodsman:** Some, but mostly lions and tigers and bears.

**Dorothy:** Lions?

**Scarecrow:** And tigers?

**Tin Woodsman:** And bears.

African cats:

- [Lions](#)
- [Leopards](#)
- [Cheetahs](#)

# Google Sets

**Tin Woodsman:** Some, but mostly lions and tigers and bears.

**Dorothy:** Lions?

**Scarecrow:** And tigers?

**Tin Woodsman:** And bears.

African cats:

- [Lions](#)
- [Leopards](#)
- [Cheetahs](#)

**Animal Toys (Ages 2-6):**

- Bears
- Puppies
- Dolphins
- Wooden

# Google Sets

**Tin Woodsman:** Some, but mostly lions and tigers and bears.

**Dorothy:** Lions?

**Scarecrow:** And tigers?

**Tin Woodsman:** And bears.

**African cats:**

- [Lions](#)
- [Leopards](#)
- [Cheetahs](#)

10:03:37 [cheetah pics]

10:05:51 [leopard pics]

**Animal Toys (Ages 2-6):**

- Bears
- Puppies
- Dolphins
- Wooden

# Google Sets

**Tin Woodsman:** Some, but mostly lions and tigers and bears.

**Dorothy:** Lions?

**Scarecrow:** And tigers?

**Tin Woodsman:** And bears.

African cats:

- [Lions](#)
- [Leopards](#)

10:03:37 [cheetah pics]

10:05:51 [leopard pics]

[U.S. Senate Committee on Environment and Public Works](#)

Exotic animals, **such as lions**, tigers, servals, monkeys, bears, snakes, iguanas, wolves, prairie dogs, and binturongs are being privately possessed as ...

[epw.senate.gov/hearing\\_statements.cfm?id=213174](http://epw.senate.gov/hearing_statements.cfm?id=213174) - 13k -

es 2-6):

# Statistical Machine Translation

- Collect parallel texts

SEHR GEEHRTER GAST!  
KUNST, KULTUR UND  
KOMFORT IM HERZEN  
BERLIN.

DEAR GUESTS,  
ART, CULTURE AND  
LUXURY IN THE HEART  
OF BERLIN.

DIE ÖRTLICHE  
NETZSPANNUNG  
BETRÄGT 220/240 VOLT  
BEI 50 HERTZ.

THE LOCAL VOLTAGE  
IS 220/240 VOLTS 50 HZ.

DE

EN

# Statistical Machine Translation

- Align

KUNST, KULTUR UND KOMFORT IM HERZEN BERLINS.



ART, CULTURE AND LUXURY IN THE HEART OF BERLIN.

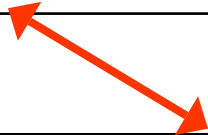


# Statistical Machine Translation

- Align

KUNST, KULTUR UND KOMFORT IM HERZEN BERLINS.

ART, CULTURE AND LUXURY IN THE HEART OF BERLIN.

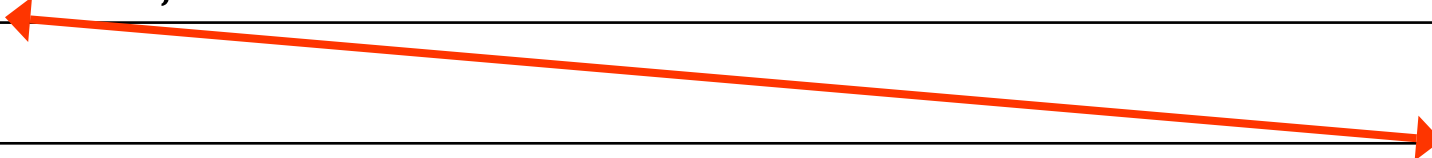


# Statistical Machine Translation

- Align

KUNST, KULTUR UND KOMFORT IM HERZEN BERLINS.

ART, CULTURE AND LUXURY IN THE HEART OF BERLIN.

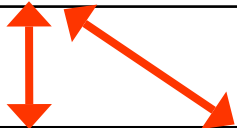


# Statistical Machine Translation

- Align

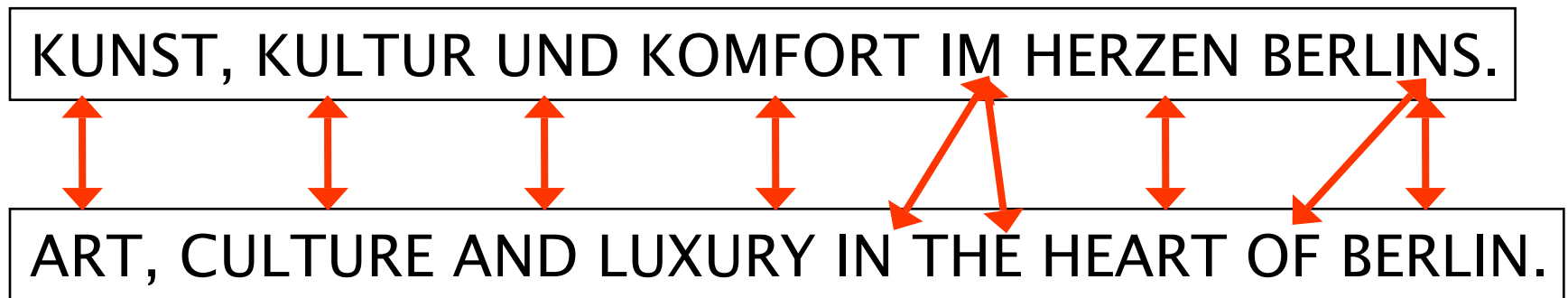
KUNST, KULTUR UND KOMFORT IM HERZEN BERLINS.

ART, CULTURE AND LUXURY IN THE HEART OF BERLIN.



# Statistical Machine Translation

- Align



# Statistical Machine Translation

- الشمالية كوريا ضد عقوبات فرض المخول الوحيد هو الامن مجلس ان يذكر حرب اعلان بمثابة العقوبات ستعتبر انها من حذرت التي.
- التركي الجيش اجتاح عندما 1974 العام منذ شطرين الى مقسمة وقبرص ضم بهدف قبارصة قوميون نفذه انقلاب على ردا الجزيرة من الشمالى الثلث اليونان الى الجزيرة.
- سنوات ثمانى استمرت العراق ضد حربا خاضت التي ايران ان يذكر تتهمه الذي العراق على اميركى عسكري هجوم شن تعارض (1980-1988) بالارهاب مرتبط وبانه الشامل للدمار اسلحة بامتلاك واشنطن.

# Statistical Machine Translation

- It is noteworthy that the Security Council is the only authorized to impose sanctions against North Korea, which warned that it would consider sanctions a declaration of war.
- Cyprus has been divided into two parts since the year 1974 when the Turkish army invaded the northern third of the island in response to a coup by Greek nationalists with the aim of annexing the island to Greece.
- It is worth mentioning that Iran, which fought a war against Iraq lasted eight years (1980 – 1988) opposes American military attack on Iraq, which Washington accuses of possessing weapons of mass destruction and that it was linked to terrorism.

# Statistical Machine Translation

- 新华社大马士革4月15日电(记者拱振喜)叙利亚总统巴沙尔·阿萨德15日在此间与来访的美国国务卿鲍威尔举行了会谈, 双方讨论了中东局势的最新发展, 特别是巴勒斯坦的严重局势以及黎以边界地区的紧张局势等问题。
- 叙利亚通讯社报道, 巴沙尔总统在会谈中说:“在巴勒斯坦发生的事件使(中东)和平进程走进了死胡同, 如果不能认识到这一点, 事情的发展有可能达到无法挽回的程度, 那时, 我们只能再等待一代人的时间。”
- 他指出, 只有在以色列从它占领的巴勒斯坦领土撤军, 停止屠杀巴勒斯坦人以后, 才可以谈和平进程的问题。

# Statistical Machine Translation

- Xinhua News Agency, Damascus, April 15 (Reporter Gong Zhenxi) Syrian President Bashar Assad 15th here with visiting US Secretary of State Colin Powell held talks, the two sides discussed the latest development of the situation in the Middle East, especially the serious situation in Palestine and the tension in the border region between Lebanon and Israel and other issues.
- According to the Syrian News Agency, President Bashar during the talks, said: "In the incident to the Palestinian (Middle East) peace process into a dead end, if not realize that this is happening may not be able to restore to the extent that time, we can only wait for the generation of time."
- He pointed out that only in the Israeli withdrawal from the occupied Palestinian territories, stop massacre of Palestinians, can talk about the peace process.





他信也说自己仍然是总理，拒绝辞职。

他信也说自己仍然是总理，拒绝辞职。

他

he

He

other

his

him

other

that he

he was

him to

he is

he has

of his

他信也说自己仍然是总理，拒绝辞职。

他

he

He

other

his

him

other

that he

he was

him to

he is

he has

of his

信

letter

trust

letters

believe

signal

a letter

believe that

letter of

confidence

他信也说自己仍然是总理，拒绝辞职。

他

信

他信

Thaksin

Thaksin Chinnawat

and Joint Communique

Dr Thaksin

Joint Communique

, Mr Thaksin

in his letter

his letter

others

he has

of his

他信也说自己仍然是总理，拒绝辞职。

他

h 他信

E Thaksin

O Thaksin Chinawatt

h and Joint Communique

h Dr Thaksin 他信也

O Joint Co: Thaksin also

tl , Mr Thaksin

h in his letter

h his letter

h others

he has

of his

他信也说自己仍然是总理，拒绝辞职。

他

他信

Thaksin

Thaksin Chinnawat  
and Joint Communique

Dr Thaksin 他信也

Joint Co: Thaksin also

, Mr Thaksin

in his letter

his letter

others

he has

of his

说自己

themselves

said that

say they

said he

say that

said they

themselves

saying that he

would say that

said that she had

saying that he has

他信也说自己仍然是总理，拒绝辞职。

他

他信

Thaksin

Thaksin Chinnawat  
and Joint Communique

Dr Thaksin 他信也

Joint Co: Thaksin also

, Mr Thaksin

in his letter

his letter

others

he has

of his

说自己

themse

said tha

say the

said he

say tha

said the

themse

saying

would s

said that she had

saying that he has

仍然是

continues to be

are still the main

would still be

continued to be

remains one of

remains one

continues to be the

still is

remains an area

still viewed by

are always one of



他信也说自己仍然是总理，拒绝辞职。

他

他信

Thaksin

Thaksin Chinnawat

and Joint Communique

Dr Thaksin 他信也

Joint Co: Thaksin also

, Mr Thaksin

in his letter

his letter

others

he has

of his

说自己

themse

said the

say the

said he

say that

said the

themse

saying

would s

said that

saying that

仍然是

contin

are sti

would

contin

remains

remains one

continues to be the

still is

remains an area

still viewed by

are always one of

said that she had

saying that he has

是总理

Prime Minister

the Prime Minister

is the Prime Minister

他信也说自己仍然是总理，拒绝辞职。

他

他信

Thaksin

Thaksin Chinnawat

and Joint Communique

Dr Thaksin 他信也

Joint Co: Thaksin also

, Mr Thaksin

in his letter

his letter

others

he has

of his

说自己

themse

said the

say the

said he

say that

said the

themse

saying

would s

said that

saying that

仍然是

contin

are sti

would

contin

remains

remains

contin

still is

remains

still

are

said

saying

是总理

Prime Minister

the Prime Minister

is the Prime Minister

总理，拒绝

.....

他信也说自己仍然是总理，拒绝辞职。

他

他信

Thaksin  
Thaksin Chinnawat  
and Joint Communique

Dr Thaksin  
Joint Co: Thaksin also

, Mr Thaksin  
in his letter  
his letter  
others

he has  
of his

说自己

themse  
said the  
say the  
said he  
say that  
said the  
themse  
saying  
would s  
said that she had  
saying that he has

仍然是

are sti  
would  
contin  
remains  
remains  
continue  
still is  
remains  
still view  
are alway

是总理

Prime Minister  
the Prime Minister  
is the Prime Minister

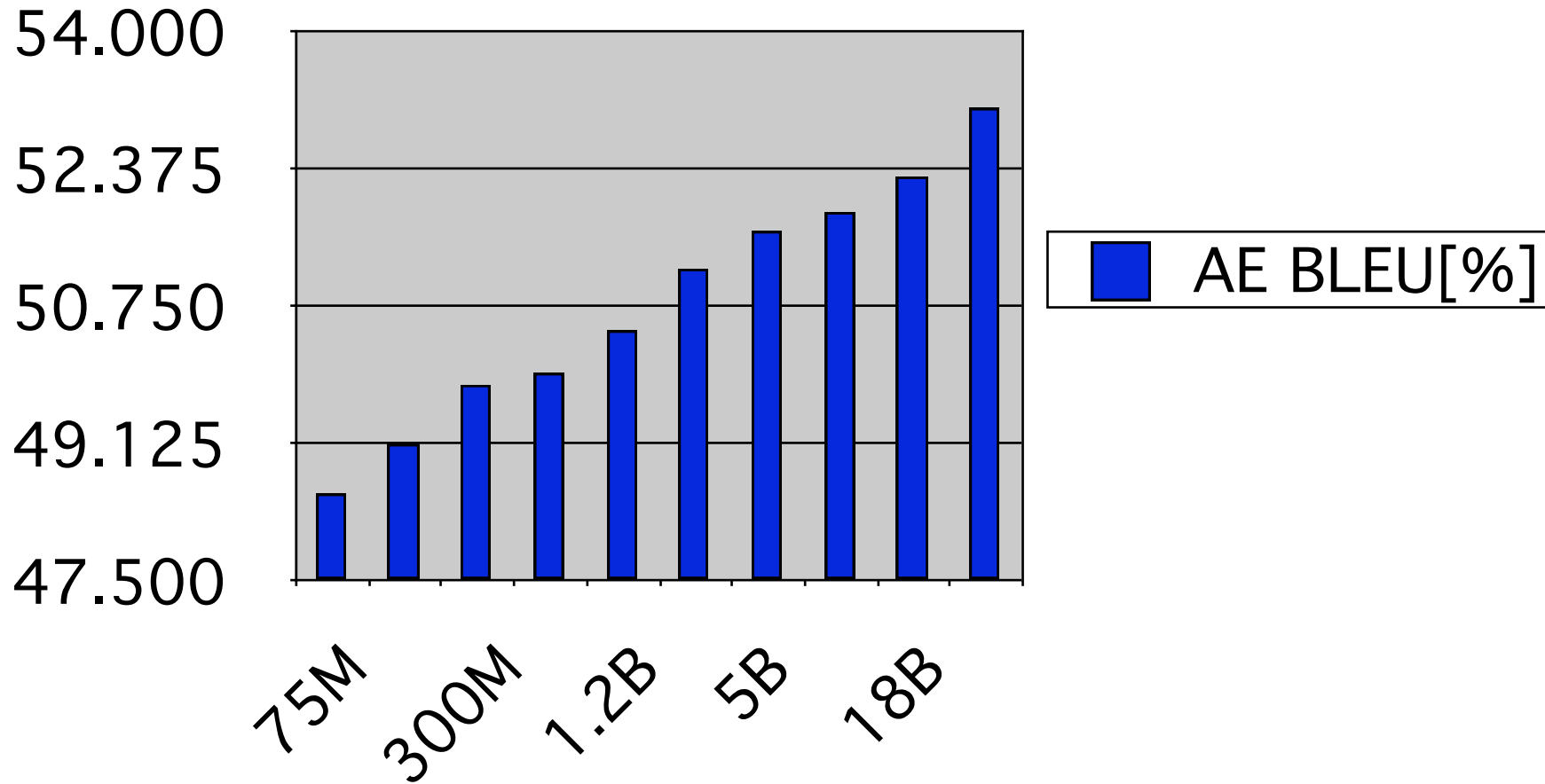
总理，拒绝

....

辞职。

resign .  
leaving their service .  
of leaving their service .  
resigned as counsel .

# More Data Still Helps



A photograph of three zebras in a savanna setting. The zebras are positioned in a row, facing slightly to the left. The background is a soft-focus landscape of tall grasses under a bright sky. The text "Image Data" is overlaid in a large, bold, black font across the center of the image.

# Image Data

# Jing, Baluja, Rowley, Google: Finding Canonical Images

[Web](#) [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more ▾](#)

[Sign in](#)



[Advanced Image Search Preferences](#)

Strict SafeSearch is on

[New! Google Image Labeler](#)

Images Showing:

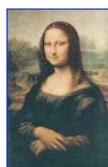
Results 1 - 21 of about 343,000 for mona lisa with Safesearch on. (0.04 seconds)



Word has it that **Mona Lisa** wasn't a ...  
320 x 366 - 21k - jpg  
[uk.gizmodo.com](http://uk.gizmodo.com)



da Vinci: **Mona Lisa**  
340 x 472 - 10k - gif  
[www.enchantedlearning.com](http://www.enchantedlearning.com)



**Mona Lisa** We have examined the topic ...  
379 x 589 - 63k - jpg  
[thesituationist.wordpress.com](http://thesituationist.wordpress.com)



**Mona Lisa** right  
282 x 795 - 59k - jpg  
[www.museumdiv.com](http://www.museumdiv.com)



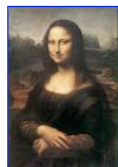
**Mona Lisa** made from train tickets --  
468 x 296 - 67k - jpg  
[www.pinktentacle.com](http://www.pinktentacle.com)



Image: **MonaLisa** sfumato.jpeg  
350 x 400 - 26k - jpeg  
[commons.wikimedia.org](http://commons.wikimedia.org)



Image: **Mona Lisa**.jpg  
743 x 1155 - 156k - jpg  
[commons.wikimedia.org](http://commons.wikimedia.org)



**MonaLisa**.jpg  
435 x 644 - 43k - jpg  
[www.mentalfloss.com](http://www.mentalfloss.com)



Study Page: **Mona Lisa** in Book Cover ...  
360 x 595 - 85k - gif  
[www.studiolo.org](http://www.studiolo.org)



**Mona Lisa**  
406 x 302 - 46k - jpg  
[www.sunrise-divers.com](http://www.sunrise-divers.com)



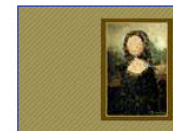
**mona lisa**  
400 x 612 - 48k - jpg  
[www.whyytraveltofrance.com](http://www.whyytraveltofrance.com)



**Mona Lisa** cartoon 3 - catalog ...  
400 x 395 - 51k - jpg  
[www.cartoonstock.com](http://www.cartoonstock.com)



**Mona Lisa** cartoon 4 - catalog ...  
400 x 400 - 51k - jpg  
[www.cartoonstock.com](http://www.cartoonstock.com)



**Mona Lisa**  
800 x 600 - 97k - jpg  
[www.vladstudio.com](http://www.vladstudio.com)



**Mona Lisa** - Joint Poster  
299 x 450 - 42k - jpg  
[www.allposters.com](http://www.allposters.com)



"**Mona Lisa**"  
507 x 694 - 22k - jpg  
[www.oregoncoastradio.com](http://www.oregoncoastradio.com)



**Mona Lisa** is Lisa Gherardini  
334 x 520 - 17k - jpg  
[yedda.com](http://yedda.com)



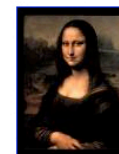
Click here if your browser does not ...  
605 x 790 - 187k - jpg  
[www.paris.org](http://www.paris.org)



Sir Joshua's **Mona Lisa**  
502 x 502 - 50k - jpg  
[www.moviespring.com](http://www.moviespring.com)



Complete history of **Mona Lisa**  
450 x 328 - 22k - jpg  
[www.simplonpc.co.uk](http://www.simplonpc.co.uk)

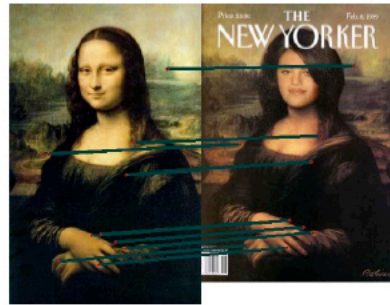


**Mona Lisa** Magnet by Leonardo da ...  
348 x 450 - 29k - jpg  
[www.allposters.com](http://www.allposters.com)

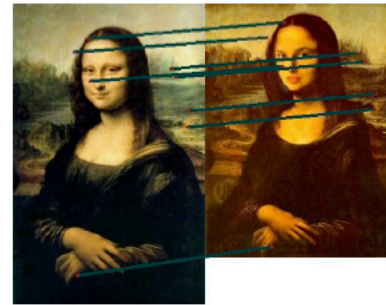
Go o o o o o o o o o o o o g l e   
1 2 3 4 5 6 7 8 9 10 [Next](#)

[New!](#) Want to help improve Google Image Search? Try [Google Image Labeler](#).

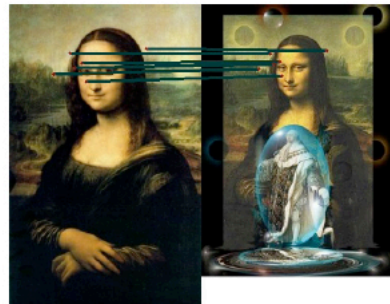
# Compare low-level features



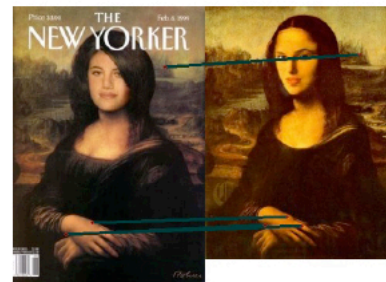
(a) A v.s. B



(b) A v.s. C



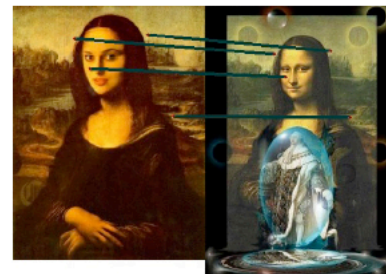
(c) A v.s. D



(d) B v.s. C

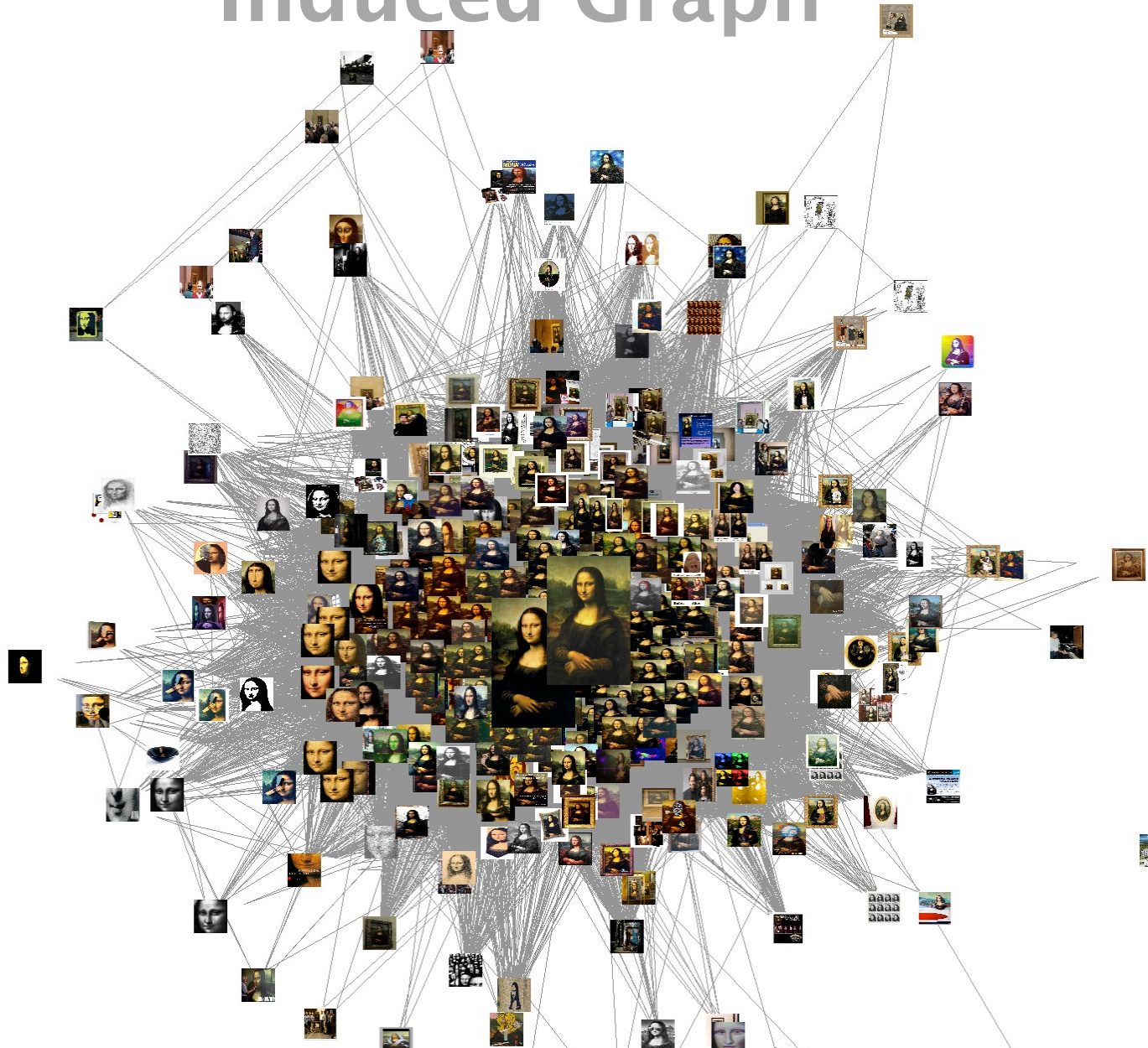


(e) B v.s. D

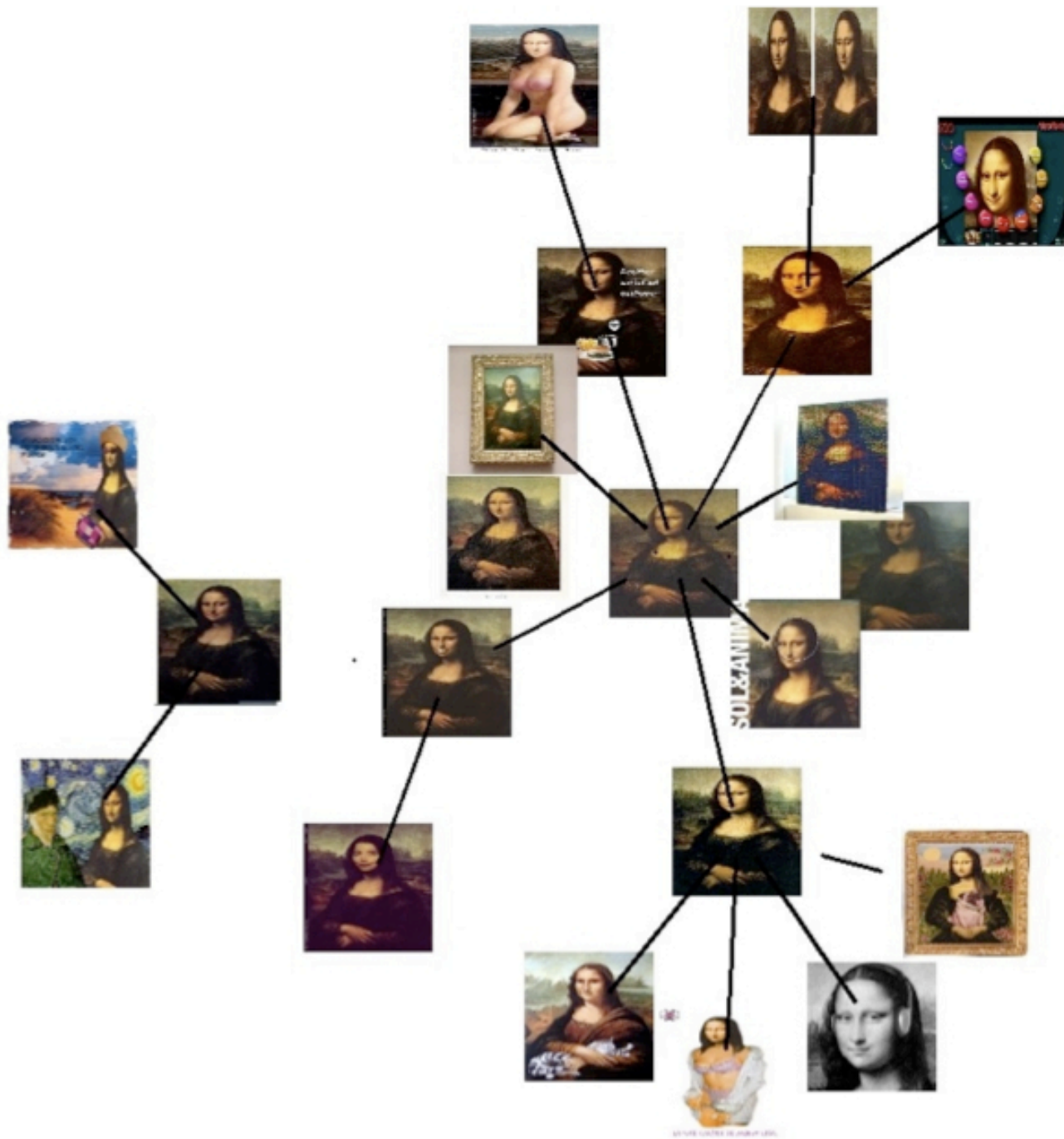


(f) C v.s. D

# Induced Graph

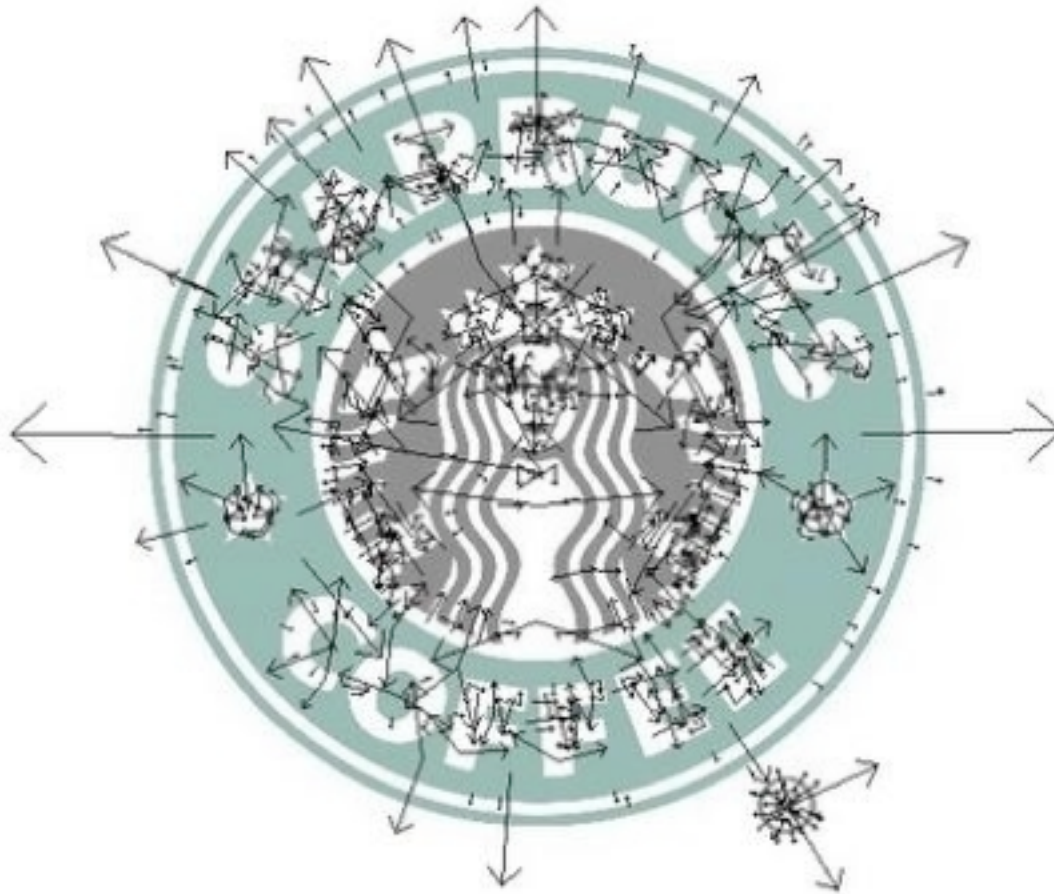


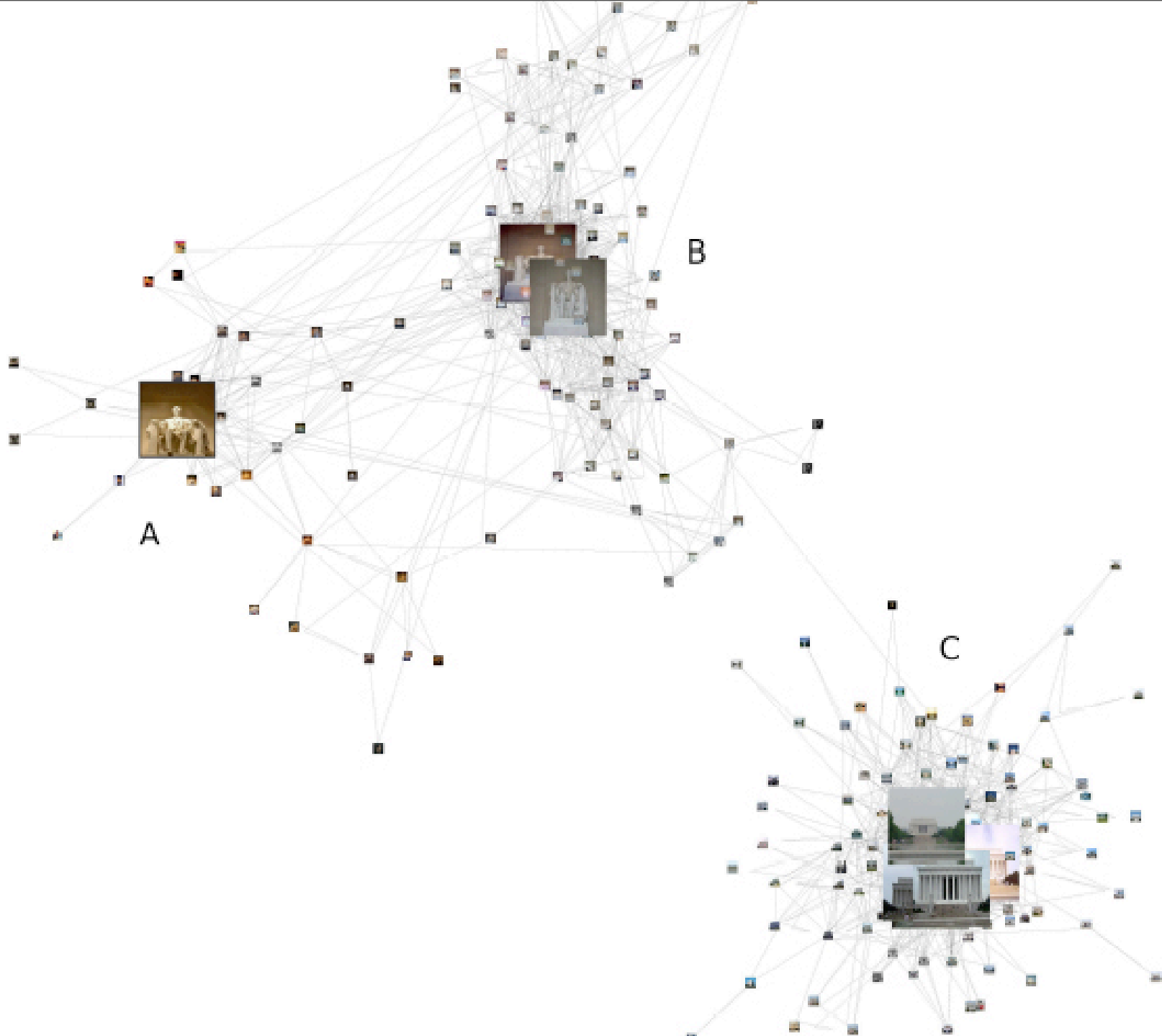






# SIFT Features





# Jay Yagnik, Atiq Islam, Google: Learning People Annotation

# Jay Yagnik, Atiq Islam, Google: Learning People Annotation

By [Guardian Unlimited](#) / [Europe](#) / [USA](#) July 13 11:33



George Bush, Angela Merkel and a barrel of Bismarck herrings. Photograph: Heribert Proepper/AP

Here's a strange set of things that come together more often than you would think: George Bush, Germany and fish. A few

# Jay Yagnik, Atiq Islam, Google: Learning People Annotation

George Bush



Alan Alda



# In Conclusion



# Code is a Liability

*“Measuring programming progress by lines of code is like measuring aircraft building progress by weight.”*

- Bill Gates

# Data is the Ultimate Asset

- Test-Driven Development
- Easy to generate a new version
- Compositional
- Results oriented
- Simple
- Easy to update for new circumstances
- Works with or without understanding
- Faster **and** better (in many cases)

**Questions?**