

CC5212-1

PROCESAMIENTO MASIVO DE DATOS
OTOÑO 2018

Lecture 7

Information Retrieval: Ranking

Aidan Hogan
aidhog@gmail.com

Apache Lucene



- Inverted Index
 - They built one so you don't have to!
 - Open Source in Java

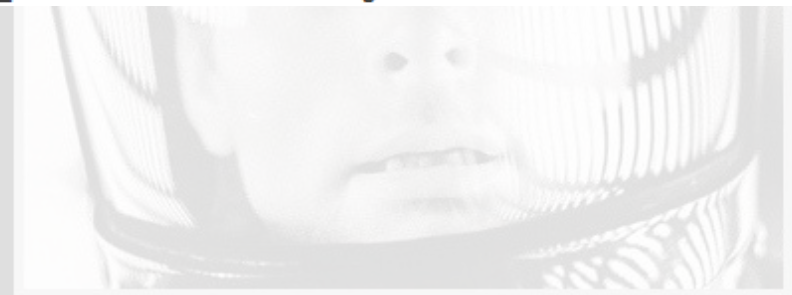


My God. It's full of win.

Apache Lucene



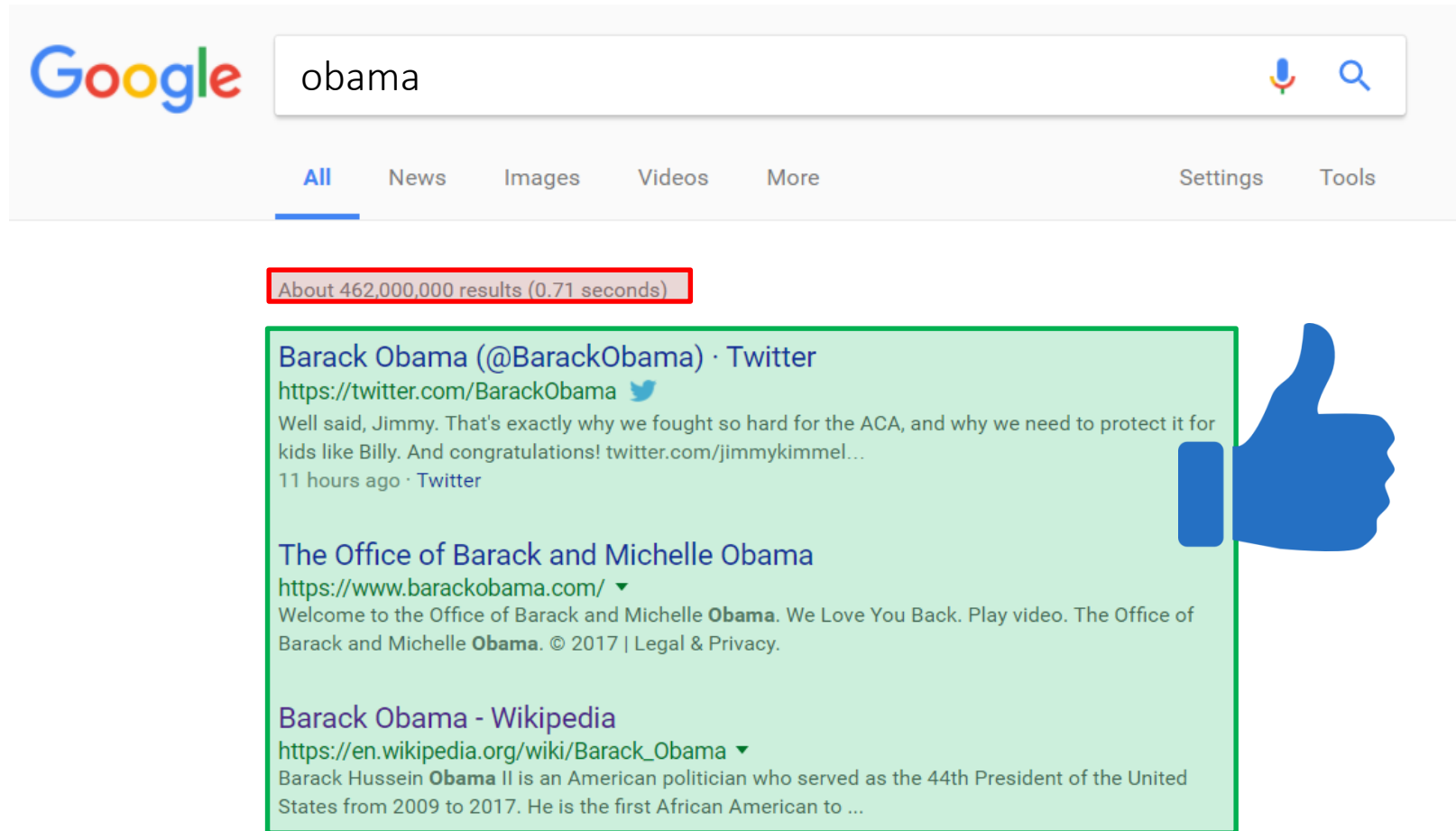
```
Tasks Console
SearchWikiIndex [Java Application] C:\Program Files\Java\jre1.8.0_65\bin\javaw.exe (03-05-2017 12:45:22 a. m.)
Opening directory at lucene
Enter a keyword search phrase:
obama
Running query: obama
Parsed query: TITLE:obam^5.0 ABSTRACT:obam
Matching documents: 255
Showing top 10 results
1 http://es.wikipedia.org/wiki/Obama_Republican Obama Republican
2 http://es.wikipedia.org/wiki/Obama_(Fukui) Obama (Fukui)
3 http://es.wikipedia.org/wiki/Republicanos_por_Obama Republicanos por Obama
4 http://es.wikipedia.org/wiki/Engonga_Obame Engonga Obame
5 http://es.wikipedia.org/wiki/Barack_Obama Barack Obama
6 http://es.wikipedia.org/wiki/Michelle_Obama Michelle Obama
7 http://es.wikipedia.org/wiki/Cartel_%22Hope%22_de_Obama Cartel "Hope" de Obama
8 http://es.wikipedia.org/wiki/Transición_presidencial_de_Barack_Obama Transición presidencial de Barack Obama
9 http://es.wikipedia.org/wiki/Por_qué_Obama_ganará_en_2008_y_en_2012 Por qué Obama ganará en 2008 y en 2012
10 http://es.wikipedia.org/wiki/Ricardo_Mangue_Obama_Nfubea Ricardo Mangue Obama Nfubea
```



My God. It's full of win.

INFORMATION RETRIEVAL: RANKING

How Does Google Get Such Good Results?




The image is a screenshot of a Google search interface. At the top left is the Google logo. To its right is a search bar containing the text 'obama'. Further right in the search bar are icons for voice search and a magnifying glass. Below the search bar is a horizontal menu with tabs: 'All' (which is underlined in blue), 'News', 'Images', 'Videos', 'More', 'Settings', and 'Tools'. Below the menu, a red-bordered box contains the text 'About 462,000,000 results (0.71 seconds)'. Below this, a green-bordered box contains three search results. To the right of the green box is a large blue thumbs-up icon. The first result is for Barack Obama's Twitter profile, the second is for the official website of the Obama administration, and the third is the Wikipedia entry for Barack Obama.

Google

obama

All News Images Videos More Settings Tools

About 462,000,000 results (0.71 seconds)

Barack Obama (@BarackObama) · Twitter
<https://twitter.com/BarackObama> 
Well said, Jimmy. That's exactly why we fought so hard for the ACA, and why we need to protect it for kids like Billy. And congratulations! [twitter.com/jimmykimmel...](https://twitter.com/jimmykimmel)
11 hours ago · [Twitter](#)

The Office of Barack and Michelle Obama
<https://www.barackobama.com/> ▼
Welcome to the Office of Barack and Michelle **Obama**. We Love You Back. Play video. The Office of Barack and Michelle **Obama**. © 2017 | [Legal & Privacy](#).

Barack Obama - Wikipedia
https://en.wikipedia.org/wiki/Barack_Obama ▼
Barack Hussein **Obama** II is an American politician who served as the 44th President of the United States from 2009 to 2017. He is the first African American to ...

How does Google Get Such Good Results?

Google

that one movie where the guy breaks his leg and spies on his neighbor

Web Videos News Images Shopping More Search tools

About 64,700,000 results (0.91 seconds)

Rear Window (1954) - IMDb
www.imdb.com/title/tt0047396/ - Internet Movie Database
★★★★★ Rating: 8.6/10 - 274,497 votes

Google

da da da dum symphony

Web Videos News Shopping Images More Search tools

About 107,000 results (0.36 seconds)



Beethoven - Symphony No. 5 in C Minor (1) - YouTube
www.youtube.com/watch?v=W2qW6fOtAMY



sometimes when i'm

- sometimes when i'm **alone i use comic sans**
- sometimes when i'm **alone i google myself**
- sometimes when i'm **alone i cry**
- sometimes when i'm **all alone**
- sometimes when i'm **dreaming**
- sometimes when i'm **sad i like to cut myself**
- sometimes when i'm **dreaming lyrics**
- sometimes when i'm **alone**
- sometimes when i'm **driving on the road at night**
- sometimes when i'm **alone i wonder**

Google Search I'm Feeling Lucky

Two Sides to Ranking: Relevance



The image shows a Google search interface for the term "obama". The search bar contains the text "obama". Below the search bar, the "Web" tab is selected, showing search results. The results include a Wikipedia entry for "Broccoli" and an article from "The World's Healthiest Foods" about broccoli's health benefits. A large red "X" is drawn over the search results, indicating a lack of relevance. In the bottom left corner, there is a cutout image of Barack Obama's head and shoulders, smiling. In the bottom right corner, there is a cutout image of a head of broccoli. The overall message is that the search results for "obama" are not relevant to the image of Barack Obama, despite the word "obama" being in the search bar.

Google

obama

Web Images News Videos More ▾ Search tools

About 16,700,000 results (0.23 seconds)

Broccoli - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Broccoli** ▾
Broccoli is an edible green plant in the cabbage family, whose large flowering head is used as a vegetable. The word **broccoli** comes from the Italian plural of ...
Cauliflower - Romanesco broccoli - Broccoli (disambiguation) - Broccolini

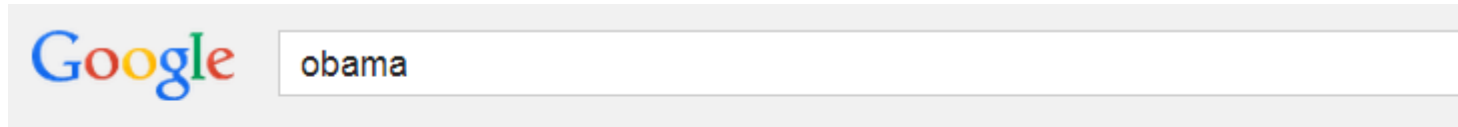
Broccoli - The World's Healthiest Foods
www.whfoods.com/genpage.php?tname=foodspice&dbid=9 ▾
Broccoli can provide you with some special cholesterol-lowering benefits if you will cook it by steaming. The fiber-related components in **broccoli** do a better job ...

News for broccoli

Mistakes We All Make With Spaghetti, Steak And ...
Huffington Post - 2 days ago
But in her new book Brassicas: Cooking the World's Healthiest Vegetables, she sa...
plunking **broccoli**, cauliflower or Brussels sprouts into ...



Two Sides to Ranking: Importance



Web Images News Videos More ▾ Search tools

About 48,100,000 results (0.26 seconds)

Mount Obama - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Mount_Obama ▾

Mount Obama (known as **Boggy Peak** until August 4, 2009) is the highest point in the nation of Antigua and Barbuda and on the island of Antigua. It lies in the far ...

Images for mount obama

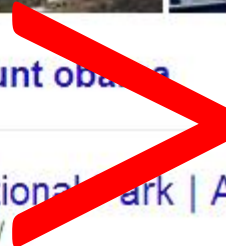
[Report images](#)



[More images for mount obama](#)

Mount Obama National Park | Antigua and Barbuda
antiguamountobama.com/

Jun 16, 2011 - As the **Mount Obama** Committee continues its work in the Area, the committee organized a site visit to the O



RANKING:
RELEVANCE

Example Query

Which of these three keyword terms is most “important”?



movie freedom wallace

Web

Images

News

Videos

More ▼

Search tools

About 4,290,000 results (0.29 seconds)

Braveheart In Defiance Of The English Tyranny! BRAVO ...



www.youtube.com/watch?v=WLrrBs8JBQo ▼

Feb 25, 2008 - Uploaded by popthetime

... actor starring as the "William **Wallace**" character in the **movie** - B...

... Braveheart **Freedom** Speech (HD) by ...

Braveheart - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Braveheart ▼

Braveheart is a 1995 epic historical drama war **film** directed by and starring Mel Gibson.

Gibson portrays ... **Wallace** instead shouts the word "**Freedom!**" and the ...

Braveheart (1995) - Quotes - IMDb

www.imdb.com/title/tt0112573/quotes ▼

... (1995) Quotes on IMDb: Memorable quotes and exchanges from **movies**, TV series and more... ... William **Wallace**: It's all for nothing if you don't have **freedom**.

Matches in a Document

W Braveheart - Wikipedia

Es seguro | https://en.wikipedia.org/wiki/Braveheart

Not logged in | **freedom** | 1 de 7

Article Talk Read Edit View history Search Wikipedia

Braveheart

From Wikipedia, the free encyclopedia

For other uses, see [Braveheart \(disambiguation\)](#).

Braveheart is a 1995 American [epic war film](#) directed by and starring [Mel Gibson](#). Gibson portrays [William Wallace](#), a 13th-century Scottish warrior who led the Scots in the [First War of Scottish Independence](#) against King Edward I of England. The story is inspired by [Blind Harry's](#) epic poem *The Actes and Deidis of the Illustre and Vallyeant Campioun Schir William Wallace* and was adapted for the screen by [Randall Wallace](#).

The film was nominated for ten [Academy Awards](#) at the 68th Academy Awards and won five: [Best Picture](#), [Best Director](#), [Best Cinematography](#), [Best Makeup](#), and [Best Sound Editing](#).

Contents [hide]

freedom
• 7 occurrences

1 Plot
2 Cast
3 Production

Upload file

Braveheart
MEL GIBSON
Every man dies,
not every man
really lives.
BRAVEHEART

Matches in a Document

W Braveheart - Wikipedia

Es seguro | https://en.wikipedia.org/wiki/Braveheart

Not logged in | movie | 3 de 16

Article Talk Read Edit View history Search Wikipedia

Braveheart

From Wikipedia, the free encyclopedia

For other uses, see [Braveheart \(disambiguation\)](#).

Braveheart is a 1995 American [epic war film](#) directed by and starring [Mel Gibson](#). Gibson portrays [William Wallace](#), a 13th-century Scottish warrior who led the Scots in the [First War of Scottish Independence](#) against King [Edward I of England](#). The story is inspired by [Blind Harry's](#) epic poem *The Actes and Deidis of the Illustre and Vallyeant Campioun Schir William Wallace* and was adapted for the screen by [Randall Wallace](#).

The film was nominated for ten [Academy Awards](#) at the 68th Academy Awards and won five: [Best Picture](#), [Best Director](#), [Best Cinematography](#), [Best Makeup](#), and [Best Sound Editing](#).

Contents [hide]

1 Plot

2 Cast

3 Production

4 Reception

5 Trivia

6 External links

7 References

8 Notes

9 Citations

10 Further reading

11 External links

12 See also

13 Images

14 Audio

15 Video

16 Quizzes

17 Games

18 Merchandise

19 Awards and nominations

20 Trivia

21 External links

22 References

23 Notes

24 Citations

25 Further reading

26 External links

27 See also

28 Images

29 Audio

30 Video

31 Quizzes

32 Games

33 Merchandise

34 Awards and nominations

35 Trivia

36 External links

37 References

38 Notes

39 Citations

40 Further reading

41 External links

42 See also

43 Images

44 Audio

45 Video

46 Quizzes

47 Games

48 Merchandise

49 Awards and nominations

50 Trivia

51 External links

52 References

53 Notes

54 Citations

55 Further reading

56 External links

57 See also

58 Images

59 Audio

60 Video

61 Quizzes

62 Games

63 Merchandise

64 Awards and nominations

65 Trivia

66 External links

67 References

68 Notes

69 Citations

70 Further reading

71 External links

72 See also

73 Images

74 Audio

75 Video

76 Quizzes

77 Games

78 Merchandise

79 Awards and nominations

80 Trivia

81 External links

82 References

83 Notes

84 Citations

85 Further reading

86 External links

87 See also

88 Images

89 Audio

90 Video

91 Quizzes

92 Games

93 Merchandise

94 Awards and nominations

95 Trivia

96 External links

97 References

98 Notes

99 Citations

100 Further reading

101 External links

102 See also

103 Images

104 Audio

105 Video

106 Quizzes

107 Games

108 Merchandise

109 Awards and nominations

110 Trivia

111 External links

112 References

113 Notes

114 Citations

115 Further reading

116 External links

117 See also

118 Images

119 Audio

120 Video

121 Quizzes

122 Games

123 Merchandise

124 Awards and nominations

125 Trivia

126 External links

127 References

128 Notes

129 Citations

130 Further reading

131 External links

132 See also

133 Images

134 Audio

135 Video

136 Quizzes

137 Games

138 Merchandise

139 Awards and nominations

140 Trivia

141 External links

142 References

143 Notes

144 Citations

145 Further reading

146 External links

147 See also

148 Images

149 Audio

150 Video

151 Quizzes

152 Games

153 Merchandise

154 Awards and nominations

155 Trivia

156 External links

157 References

158 Notes

159 Citations

160 Further reading

161 External links

162 See also

163 Images

164 Audio

165 Video

166 Quizzes

167 Games

168 Merchandise

169 Awards and nominations

170 Trivia

171 External links

172 References

173 Notes

174 Citations

175 Further reading

176 External links

177 See also

178 Images

179 Audio

180 Video

181 Quizzes

182 Games

183 Merchandise

184 Awards and nominations

185 Trivia

186 External links

187 References

188 Notes

189 Citations

190 Further reading

191 External links

192 See also

193 Images

194 Audio

195 Video

196 Quizzes

197 Games

198 Merchandise

199 Awards and nominations

200 Trivia

201 External links

202 References

203 Notes

204 Citations

205 Further reading

206 External links

207 See also

208 Images

209 Audio

210 Video

211 Quizzes

212 Games

213 Merchandise

214 Awards and nominations

215 Trivia

216 External links

217 References

218 Notes

219 Citations

220 Further reading

221 External links

222 See also

223 Images

224 Audio

225 Video

226 Quizzes

227 Games

228 Merchandise

229 Awards and nominations

230 Trivia

231 External links

232 References

233 Notes

234 Citations

235 Further reading

236 External links

237 See also

238 Images

239 Audio

240 Video

241 Quizzes

242 Games

243 Merchandise

244 Awards and nominations

245 Trivia

246 External links

247 References

248 Notes

249 Citations

250 Further reading

251 External links

252 See also

253 Images

254 Audio

255 Video

256 Quizzes

257 Games

258 Merchandise

259 Awards and nominations

260 Trivia

261 External links

262 References

263 Notes

264 Citations

265 Further reading

266 External links

267 See also

268 Images

269 Audio

270 Video

271 Quizzes

272 Games

273 Merchandise

274 Awards and nominations

275 Trivia

276 External links

277 References

278 Notes

279 Citations

280 Further reading

281 External links

282 See also

283 Images

284 Audio

285 Video

286 Quizzes

287 Games

288 Merchandise

289 Awards and nominations

290 Trivia

291 External links

292 References

293 Notes

294 Citations

295 Further reading

296 External links

297 See also

298 Images

299 Audio

300 Video

301 Quizzes

302 Games

303 Merchandise

304 Awards and nominations

305 Trivia

306 External links

307 References

308 Notes

309 Citations

310 Further reading

311 External links

312 See also

313 Images

314 Audio

315 Video

316 Quizzes

317 Games

318 Merchandise

319 Awards and nominations

320 Trivia

321 External links

322 References

323 Notes

324 Citations

325 Further reading

326 External links

327 See also

328 Images

329 Audio

330 Video

331 Quizzes

332 Games

333 Merchandise

334 Awards and nominations

335 Trivia

336 External links

337 References

338 Notes

339 Citations

340 Further reading

341 External links

342 See also

343 Images

344 Audio

345 Video

346 Quizzes

347 Games

348 Merchandise

349 Awards and nominations

350 Trivia

351 External links

352 References

353 Notes

354 Citations

355 Further reading

356 External links

357 See also

358 Images

359 Audio

360 Video

361 Quizzes

362 Games

363 Merchandise

364 Awards and nominations

365 Trivia

366 External links

367 References

368 Notes

369 Citations

370 Further reading

371 External links

372 See also

373 Images

374 Audio

375 Video

376 Quizzes

377 Games

378 Merchandise

379 Awards and nominations

380 Trivia

381 External links

382 References

383 Notes

384 Citations

385 Further reading

386 External links

387 See also

388 Images

389 Audio

390 Video

391 Quizzes

392 Games

393 Merchandise

394 Awards and nominations

395 Trivia

396 External links

397 References

398 Notes

399 Citations

400 Further reading

401 External links

402 See also

403 Images

404 Audio

405 Video

406 Quizzes

407 Games

408 Merchandise

409 Awards and nominations

410 Trivia

411 External links

412 References

413 Notes

414 Citations

415 Further reading

416 External links

417 See also

418 Images

419 Audio

420 Video

421 Quizzes

422 Games

423 Merchandise

424 Awards and nominations

425 Trivia

426 External links

427 References

428 Notes

429 Citations

430 Further reading

431 External links

432 See also

433 Images

434 Audio

435 Video

436 Quizzes

437 Games

438 Merchandise

439 Awards and nominations

440 Trivia

441 External links

442 References

443 Notes

444 Citations

445 Further reading

446 External links

447 See also

448 Images

449 Audio

450 Video

451 Quizzes

452 Games

453 Merchandise

454 Awards and nominations

455 Trivia

456 External links

457 References

458 Notes

459 Citations

460 Further reading

461 External links

462 See also

463 Images

464 Audio

465 Video

466 Quizzes

467 Games

468 Merchandise

469 Awards and nominations

470 Trivia

471 External links

472 References

473 Notes

474 Citations

475 Further reading

476 External links

477 See also

478 Images

479 Audio

480 Video

481 Quizzes

482 Games

483 Merchandise

484 Awards and nominations

485 Trivia

486 External links

487 References

488 Notes

489 Citations

490 Further reading

491 External links

492 See also

493 Images

494 Audio

495 Video

496 Quizzes

497 Games

498 Merchandise

499 Awards and nominations

500 Trivia

501 External links

502 References

503 Notes

504 Citations

505 Further reading

506 External links

507 See also

508 Images

509 Audio

510 Video

511 Quizzes

512 Games

513 Merchandise

514 Awards and nominations

515 Trivia

516 External links

517 References

518 Notes

519 Citations

520 Further reading

521 External links

522 See also

523 Images

524 Audio

525 Video

526 Quizzes

527 Games

528 Merchandise

529 Awards and nominations

530 Trivia

531 External links

532 References

533 Notes

534 Citations

535 Further reading

536 External links

537 See also

538 Images

539 Audio

540 Video

541 Quizzes

542 Games

543 Merchandise

544 Awards and nominations

545 Trivia

546 External links

547 References

548 Notes

549 Citations

550 Further reading

551 External links

552 See also

553 Images

554 Audio

555 Video

556 Quizzes

557 Games

558 Merchandise

559 Awards and nominations

560 Trivia

561 External links

562 References

563 Notes

564 Citations

565 Further reading

566 External links

567 See also

568 Images

569 Audio

570 Video

571 Quizzes

572 Games

573 Merchandise

574 Awards and nominations

575 Trivia

576 External links

577 References

578 Notes

579 Citations

580 Further reading

581 External links

582 See also

583 Images

584 Audio

585 Video

586 Quizzes

587 Games

588 Merchandise

589 Awards and nominations

590 Trivia

591 External links

592 References

593 Notes

594 Citations

595 Further reading

596 External links

597 See also

598 Images

599 Audio

600 Video

601 Quizzes

602 Games

603 Merchandise

604 Awards and nominations

605 Trivia

606 External links

607 References

608 Notes

609 Citations

610 Further reading

611 External links

612 See also

613 Images

614 Audio

615 Video

616 Quizzes

617 Games

618 Merchandise

619 Awards and nominations

620 Trivia

621 External links

622 References

623 Notes

624 Citations

625 Further reading

626 External links

627 See also

628 Images

629 Audio

630 Video

631 Quizzes

632 Games

633 Merchandise

634 Awards and nominations

635 Trivia

636 External links

637 References

638 Notes

639 Citations

640 Further reading

641 External links

642 See also

643 Images

644 Audio

645 Video

646 Quizzes

647 Games

648 Merchandise

649 Awards and nominations

650 Trivia

651 External links

652 References

653 Notes

654 Citations

655 Further reading

656 External links

657 See also

658 Images

659 Audio

660 Video

661 Quizzes

662 Games

663 Merchandise

664 Awards and nominations

665 Trivia

666 External links

667 References

668 Notes

669 Citations

670 Further reading

671 External links

672 See also

673 Images

674 Audio

675 Video

676 Quizzes

677 Games

678 Merchandise

679 Awards and nominations

680 Trivia

681 External links

682 References

683 Notes

684 Citations

685 Further reading

686 External links

687 See also

688 Images

689 Audio

690 Video

691 Quizzes

692 Games

693 Merchandise

694 Awards and nominations

695 Trivia

696 External links

697 References

698 Notes

699 Citations

700 Further reading

701 External links

702 See also

703 Images

704 Audio

705 Video

706 Quizzes

707 Games

708 Merchandise

709 Awards and nominations

710 Trivia

711 External links

712 References

713 Notes

714 Citations

715 Further reading

716 External links

717 See also

718 Images

719 Audio

720 Video

721 Quizzes

722 Games

723 Merchandise

724 Awards and nominations

725 Trivia

726 External links

727 References

728 Notes

729 Citations

730 Further reading

731 External links

732 See also

733 Images

734 Audio

735 Video

736 Quizzes

737 Games

738 Merchandise

739 Awards and nominations

740 Trivia

741 External links

742 References

743 Notes

744 Citations

745 Further reading

746 External links

747 See also

748 Images

749 Audio

750 Video

751 Quizzes

752 Games

753 Merchandise

754 Awards and nominations

755 Trivia

756 External links

757 References

758 Notes

759 Citations

760 Further reading

761 External links

762 See also

763 Images

764 Audio

765 Video

766 Quizzes

767 Games

768 Merchandise

769 Awards and nominations

770 Trivia

771 External links

772 References

773 Notes

774 Citations

775 Further reading

776 External links

777 See also

778 Images

779 Audio

780 Video

781 Quizzes

782 Games

783 Merchandise

784 Awards and nominations

785 Trivia

786 External links

787 References

788 Notes

789 Citations

790 Further reading

791 External links

792 See also

793 Images

794 Audio

795 Video

796 Quizzes

797 Games

798 Merchandise

799 Awards and nominations

800 Trivia

801 External links

802 References

803 Notes

804 Citations

805 Further reading

806 External links

807 See also

808 Images

809 Audio

810 Video

811 Quizzes

812 Games

813 Merchandise

814 Awards and nominations

815 Trivia

816 External links

817 References

818 Notes

819 Citations

820 Further reading

821 External links

822 See also

823 Images

824 Audio

825 Video

826 Quizzes

827 Games

828 Merchandise

829 Awards and nominations

830 Trivia

831 External links

832 References

833 Notes

834 Citations

835 Further reading

836 External links

837 See also

838 Images

839 Audio

840 Video

841 Quizzes

842 Games

843 Merchandise

844 Awards and nominations

845 Trivia

846 External links

847 References

848 Notes

849 Citations

850 Further reading

851 External links

852 See also

853 Images

854 Audio

855 Video

856 Quizzes

857 Games

858 Merchandise

859 Awards and nominations

860 Trivia

861 External links

862 References

863 Notes

864 Citations

865 Further reading

866 External links

867 See also

868 Images

869 Audio

870 Video

871 Quizzes

872 Games

873 Merchandise

874 Awards and nominations

875 Trivia

876 External links

877 References

878 Notes

879 Citations

880 Further reading

881 External links

882 See also

883 Images

884 Audio

885 Video

886 Quizzes

887 Games

888 Merchandise

889 Awards and nominations

890 Trivia

891 External links

892 References

893 Notes

894 Citations

895 Further reading

896 External links

897 See also

898 Images

899 Audio

900 Video

901 Quizzes

902 Games

903 Merchandise

904 Awards and nominations

905 Trivia

906 External links

907 References

908 Notes

909 Citations

910 Further reading

911 External links

912 See also

913 Images

914 Audio

915 Video

916 Quizzes

917 Games

918 Merchandise

919 Awards and nominations

920 Trivia

921 External links

922 References

923 Notes

924 Citations

925 Further reading

926 External links

927 See also

928 Images

929 Audio

930 Video

931 Quizzes

932 Games

933 Merchandise

934 Awards and nominations

935 Trivia

936 External links

937 References

938 Notes

939 Citations

940 Further reading

941 External links

942 See also

943 Images

944 Audio

945 Video

946 Quizzes

947 Games

948 Merchandise

949 Awards and nominations

950 Trivia

951 External links

952 References

953 Notes

954 Citations

955 Further reading

956 External links

957 See also

958 Images

959 Audio

960 Video

961 Quizzes

962 Games

963 Merchandise

964 Awards and nominations

965 Trivia

966 External links

967 References

968 Notes

969 Citations

970 Further reading

971 External links

972 See also

973 Images

974 Audio

975 Video

976 Quizzes

977 Games

978 Merchandise

979 Awards and nominations

980 Trivia

981 External links

982 References

983 Notes

984 Citations

985 Further reading

986 External links

987 See also

988 Images

989 Audio

990 Video

991 Quizzes

992 Games

993 Merchandise

994 Awards and nominations

995 Trivia

996 External links

997 References

998 Notes

999 Citations

1000 Further reading

1001 External links

1002 See also

1003 Images

1004 Audio

1005 Video

1006 Quizzes

1007 Games

1008 Merchandise

1009 Awards and nominations

1010 Trivia

1011 External links

1012 References

1013 Notes

1014 Citations

1015 Further reading

1016 External links

1017 See also

1018 Images

1019 Audio

1020 Video

1021 Quizzes

1022 Games

1023 Merchandise

1024 Awards and nominations

1025 Trivia

1026 External links

1027 References

1028 Notes

1029 Citations

1030 Further reading

1031 External links

1032 See also

1033 Images

1034 Audio

1035 Video

1036 Quizzes

1037 Games

1038 Merchandise

1039 Awards and nominations

1040 Trivia

1041 External links

1042 References

1043 Notes

1044 Citations

1045 Further reading

1046 External links

1047 See also

1048 Images

1049 Audio

1050 Video

1051 Quizzes

1052 Games

1053 Merchandise

1054 Awards and nominations

1055 Trivia

1056 External links

1057 References

1058 Notes

1059 Citations

1060 Further reading

1061 External links

1062 See also

1063 Images

1064 Audio

1065 Video

1066 Quizzes

1067 Games

1068 Merchandise

1069 Awards and nominations

1070 Trivia

1071 External links

1072 References

1073 Notes

1074 Citations

1075 Further reading

1076 External links

1077 See also

1078 Images

1079 Audio

1080 Video

1081 Quizzes

1082 Games

1083 Merchandise

1084 Awards and nominations

1085 Trivia

1086 External links

1087 References

1088 Notes

1089 Citations

1090 Further reading

1091 External links

1092 See also

1093 Images

1094 Audio

1095 Video

1096 Quizzes

1097 Games

1098 Merchandise

1099 Awards and nominations

1100 Trivia

1101 External links

1102 References

1103 Notes

1104 Citations

1105 Further reading

1106 External links

1107 See also

1108 Images

1109 Audio

1110 Video

1111 Quizzes

1112 Games

1113 Merchandise

1114 Awards and nominations

1115 Trivia

1116 External links

1117 References

1118 Notes

1119 Citations

1120 Further reading

1121 External links

1122 See also

1123 Images

1124 Audio

1125 Video

1126 Quizzes

1127 Games

1128 Merchandise

1129 Awards and nominations

1130 Trivia

1131 External links

1132 References

1133 Notes

1134 Citations

1135 Further reading

1136 External links

1137 See also

1138 Images

1139 Audio

1140 Video

1141 Quizzes

1142 Games

1143 Merchandise

1144 Awards and nominations

1145 Trivia

1146 External links

1147 References

1148 Notes

1149 Citations

1150 Further reading

1151 External links

1152 See also

1153 Images

1154 Audio

1155 Video

1156 Quizzes

1157 Games

1158 Merchandise

1159 Awards and nominations

1160 Trivia

1161 External links

1162 References

1163 Notes

1164 Citations

1165 Further reading

1166 External links

1167 See also

1168 Images

1169 Audio

1170 Video

1171 Quizzes

1172 Games

1173 Merchandise

1174 Awards and nominations

1175 Trivia

1176 External links

1177 References

1178 Notes

1179 Citations

1180 Further reading

1181 External links

1182 See also

1183 Images

1184 Audio

1185 Video

1186 Quizzes

1187 Games

1188 Merchandise

1189 Awards and nominations

1190 Trivia

1191 External links

1192 References

1193 Notes

1194 Citations

1195 Further reading

1196 External links

1197 See also

1198 Images

1199 Audio

1200 Video

1201 Quizzes

1202 Games

1203 Merchandise

1204 Awards and nominations

1205 Trivia

1206 External links

1207 References

1208 Notes

1209 Citations

1210 Further reading

1211 External links

1212 See also

1213 Images

1214 Audio

1215 Video

1216 Quizzes

1217 Games

1218 Merchandise

1219 Awards and nominations

1220 Trivia

1221 External links

1222 References

1223 Notes

1224 Citations

1225 Further reading

1

Matches in a Document

W Braveheart - Wikipedia

Es seguro | https://en.wikipedia.org/wiki/Braveheart

Not logged in | wallace 44 de 88

Article Talk Read Edit View history Search Wikipedia

Braveheart

From Wikipedia, the free encyclopedia

For other uses, see *Braveheart* (disambiguation).

Braveheart is a 1995 American epic war film directed by and starring Mel Gibson. Gibson portrays William Wallace, a 13th-century Scottish warrior who led the Scots in the First War of Scottish Independence against King Edward I of England. The story is inspired by Blind Harry's epic poem *The Actes and Deidis of the Illustre and Vallyeant Campioun Schir William Wallace* and was adapted for the screen by Randall Wallace.

The film was nominated for ten Academy Awards at the 68th Academy Awards and won five: Best Picture, Best Director, Best Cinematography, Best Makeup, and Best Sound Editing.

Contents [hide]

1 Plot

2 Cast

3 Production

4 Reception

5 Trivia

6 External links

7 References

8 Notes

9 See also

10 Further reading

11 Citations

12 External links

13 References

14 Notes

15 See also

16 Further reading

17 Citations

18 External links

19 References

20 Notes

21 See also

22 Further reading

23 Citations

24 External links

25 References

26 Notes

27 See also

28 Further reading

29 Citations

30 External links

31 References

32 Notes

33 See also

34 Further reading

35 Citations

36 External links

37 References

38 Notes

39 See also

40 Further reading

41 Citations

42 External links

43 References

44 Notes

45 See also

46 Further reading

47 Citations

48 External links

49 References

50 Notes

51 See also

52 Further reading

53 Citations

54 External links

55 References

56 Notes

57 See also

58 Further reading

59 Citations

60 External links

61 References

62 Notes

63 See also

64 Further reading

65 Citations

66 External links

67 References

68 Notes

69 See also

70 Further reading

71 Citations

72 External links

73 References

74 Notes

75 See also

76 Further reading

77 Citations

78 External links

79 References

80 Notes

81 See also

82 Further reading

83 Citations

84 External links

85 References

86 Notes

87 See also

88 Further reading

89 Citations

90 External links

91 References

92 Notes

93 See also

94 Further reading

95 Citations

96 External links

97 References

98 Notes

99 See also

100 Further reading

101 Citations

102 External links

103 References

104 Notes

105 See also

106 Further reading

107 Citations

108 External links

109 References

110 Notes

111 See also

112 Further reading

113 Citations

114 External links

115 References

116 Notes

117 See also

118 Further reading

119 Citations

120 External links

121 References

122 Notes

123 See also

124 Further reading

125 Citations

126 External links

127 References

128 Notes

129 See also

130 Further reading

131 Citations

132 External links

133 References

134 Notes

135 See also

136 Further reading

137 Citations

138 External links

139 References

140 Notes

141 See also

142 Further reading

143 Citations

144 External links

145 References

146 Notes

147 See also

148 Further reading

149 Citations

150 External links

151 References

152 Notes

153 See also

154 Further reading

155 Citations

156 External links

157 References

158 Notes

159 See also

160 Further reading

161 Citations

162 External links

163 References

164 Notes

165 See also

166 Further reading

167 Citations

168 External links

169 References

170 Notes

171 See also

172 Further reading

173 Citations

174 External links

175 References

176 Notes

177 See also

178 Further reading

179 Citations

180 External links

181 References

182 Notes

183 See also

184 Further reading

185 Citations

186 External links

187 References

188 Notes

189 See also

190 Further reading

191 Citations

192 External links

193 References

194 Notes

195 See also

196 Further reading

197 Citations

198 External links

199 References

200 Notes

201 See also

202 Further reading

203 Citations

204 External links

205 References

206 Notes

207 See also

208 Further reading

209 Citations

210 External links

211 References

212 Notes

213 See also

214 Further reading

215 Citations

216 External links

217 References

218 Notes

219 See also

220 Further reading

221 Citations

222 External links

223 References

224 Notes

225 See also

226 Further reading

227 Citations

228 External links

229 References

230 Notes

231 See also

232 Further reading

233 Citations

234 External links

235 References

236 Notes

237 See also

238 Further reading

239 Citations

240 External links

241 References

242 Notes

243 See also

244 Further reading

245 Citations

246 External links

247 References

248 Notes

249 See also

250 Further reading

251 Citations

252 External links

253 References

254 Notes

255 See also

256 Further reading

257 Citations

258 External links

259 References

260 Notes

261 See also

262 Further reading

263 Citations

264 External links

265 References

266 Notes

267 See also

268 Further reading

269 Citations

270 External links

271 References

272 Notes

273 See also

274 Further reading

275 Citations

276 External links

277 References

278 Notes

279 See also

280 Further reading

281 Citations

282 External links

283 References

284 Notes

285 See also

286 Further reading

287 Citations

288 External links

289 References

290 Notes

291 See also

292 Further reading

293 Citations

294 External links

295 References

296 Notes

297 See also

298 Further reading

299 Citations

300 External links

301 References

302 Notes

303 See also

304 Further reading

305 Citations

306 External links

307 References

308 Notes

309 See also

310 Further reading

311 Citations

312 External links

313 References

314 Notes

315 See also

316 Further reading

317 Citations

318 External links

319 References

320 Notes

321 See also

322 Further reading

323 Citations

324 External links

325 References

326 Notes

327 See also

328 Further reading

329 Citations

330 External links

331 References

332 Notes

333 See also

334 Further reading

335 Citations

336 External links

337 References

338 Notes

339 See also

340 Further reading

341 Citations

342 External links

343 References

344 Notes

345 See also

346 Further reading

347 Citations

348 External links

349 References

350 Notes

351 See also

352 Further reading

353 Citations

354 External links

355 References

356 Notes

357 See also

358 Further reading

359 Citations

360 External links

361 References

362 Notes

363 See also

364 Further reading

365 Citations

366 External links

367 References

368 Notes

369 See also

370 Further reading

371 Citations

372 External links

373 References

374 Notes

375 See also

376 Further reading

377 Citations

378 External links

379 References

380 Notes

381 See also

382 Further reading

383 Citations

384 External links

385 References

386 Notes

387 See also

388 Further reading

389 Citations

390 External links

391 References

392 Notes

393 See also

394 Further reading

395 Citations

396 External links

397 References

398 Notes

399 See also

400 Further reading

401 Citations

402 External links

403 References

404 Notes

405 See also

406 Further reading

407 Citations

408 External links

409 References

410 Notes

411 See also

412 Further reading

413 Citations

414 External links

415 References

416 Notes

417 See also

418 Further reading

419 Citations

420 External links

421 References

422 Notes

423 See also

424 Further reading

425 Citations

426 External links

427 References

428 Notes

429 See also

430 Further reading

431 Citations

432 External links

433 References

434 Notes

435 See also

436 Further reading

437 Citations

438 External links

439 References

440 Notes

441 See also

442 Further reading

443 Citations

444 External links

445 References

446 Notes

447 See also

448 Further reading

449 Citations

450 External links

451 References

452 Notes

453 See also

454 Further reading

455 Citations

456 External links

457 References

458 Notes

459 See also

460 Further reading

461 Citations

462 External links

463 References

464 Notes

465 See also

466 Further reading

467 Citations

468 External links

469 References

470 Notes

471 See also

472 Further reading

473 Citations

474 External links

475 References

476 Notes

477 See also

478 Further reading

479 Citations

480 External links

481 References

482 Notes

483 See also

484 Further reading

485 Citations

486 External links

487 References

488 Notes

489 See also

490 Further reading

491 Citations

492 External links

493 References

494 Notes

495 See also

496 Further reading

497 Citations

498 External links

499 References

500 Notes

501 See also

502 Further reading

503 Citations

504 External links

505 References

506 Notes

507 See also

508 Further reading

509 Citations

510 External links

511 References

512 Notes

513 See also

514 Further reading

515 Citations

516 External links

517 References

518 Notes

519 See also

520 Further reading

521 Citations

522 External links

523 References

524 Notes

525 See also

526 Further reading

527 Citations

528 External links

529 References

530 Notes

531 See also

532 Further reading

533 Citations

534 External links

535 References

536 Notes

537 See also

538 Further reading

539 Citations

540 External links

541 References

542 Notes

543 See also

544 Further reading

545 Citations

546 External links

547 References

548 Notes

549 See also

550 Further reading

551 Citations

552 External links

553 References

554 Notes

555 See also

556 Further reading

557 Citations

558 External links

559 References

560 Notes

561 See also

562 Further reading

563 Citations

564 External links

565 References

566 Notes

567 See also

568 Further reading

569 Citations

570 External links

571 References

572 Notes

573 See also

574 Further reading

575 Citations

576 External links

577 References

578 Notes

579 See also

580 Further reading

581 Citations

582 External links

583 References

584 Notes

585 See also

586 Further reading

587 Citations

588 External links

589 References

590 Notes

591 See also

592 Further reading

593 Citations

594 External links

595 References

596 Notes

597 See also

598 Further reading

599 Citations

600 External links

601 References

602 Notes

603 See also

604 Further reading

605 Citations

606 External links

607 References

608 Notes

609 See also

610 Further reading

611 Citations

612 External links

613 References

614 Notes

615 See also

616 Further reading

617 Citations

618 External links

619 References

620 Notes

621 See also

622 Further reading

623 Citations

624 External links

625 References

626 Notes

627 See also

628 Further reading

629 Citations

630 External links

631 References

632 Notes

633 See also

634 Further reading

635 Citations

636 External links

637 References

638 Notes

639 See also

640 Further reading

641 Citations

642 External links

643 References

644 Notes

645 See also

646 Further reading

647 Citations

648 External links

649 References

650 Notes

651 See also

652 Further reading

653 Citations

654 External links

655 References

656 Notes

657 See also

658 Further reading

659 Citations

660 External links

661 References

662 Notes

663 See also

664 Further reading

665 Citations

666 External links

667 References

668 Notes

669 See also

670 Further reading

671 Citations

672 External links

673 References

674 Notes

675 See also

676 Further reading

677 Citations

678 External links

679 References

680 Notes

681 See also

682 Further reading

683 Citations

684 External links

685 References

686 Notes

687 See also

688 Further reading

689 Citations

690 External links

691 References

692 Notes

693 See also

694 Further reading

695 Citations

696 External links

697 References

698 Notes

699 See also

700 Further reading

701 Citations

702 External links

703 References

704 Notes

705 See also

706 Further reading

707 Citations

708 External links

709 References

710 Notes

711 See also

712 Further reading

713 Citations

714 External links

715 References

716 Notes

717 See also

718 Further reading

719 Citations

720 External links

721 References

722 Notes

723 See also

724 Further reading

725 Citations

726 External links

727 References

728 Notes

729 See also

730 Further reading

731 Citations

732 External links

733 References

734 Notes

735 See also

736 Further reading

737 Citations

738 External links

739 References

740 Notes

741 See also

742 Further reading

743 Citations

744 External links

745 References

746 Notes

747 See also

748 Further reading

749 Citations

750 External links

751 References

752 Notes

753 See also

754 Further reading

755 Citations

756 External links

757 References

758 Notes

759 See also

760 Further reading

761 Citations

762 External links

763 References

764 Notes

765 See also

766 Further reading

767 Citations

768 External links

769 References

770 Notes

771 See also

772 Further reading

773 Citations

774 External links

775 References

776 Notes

777 See also

778 Further reading

779 Citations

780 External links

781 References

782 Notes

783 See also

784 Further reading

785 Citations

786 External links

787 References

788 Notes

789 See also

790 Further reading

791 Citations

792 External links

793 References

794 Notes

795 See also

796 Further reading

797 Citations

798 External links

799 References

800 Notes

801 See also

802 Further reading

803 Citations

804 External links

805 References

806 Notes

807 See also

808 Further reading

809 Citations

810 External links

811 References

812 Notes

813 See also

814 Further reading

815 Citations

816 External links

817 References

818 Notes

819 See also

820 Further reading

821 Citations

822 External links

823 References

824 Notes

825 See also

826 Further reading

827 Citations

828 External links

829 References

830 Notes

831 See also

832 Further reading

833 Citations

834 External links

835 References

836 Notes

837 See also

838 Further reading

839 Citations

840 External links

841 References

842 Notes

843 See also

844 Further reading

845 Citations

846 External links

847 References

848 Notes

849 See also

850 Further reading

851 Citations

852 External links

853 References

854 Notes

855 See also

856 Further reading

857 Citations

858 External links

859 References

860 Notes

861 See also

862 Further reading

863 Citations

864 External links

865 References

866 Notes

867 See also

868 Further reading

869 Citations

870 External links

871 References

872 Notes

873 See also

874 Further reading

875 Citations

876 External links

877 References

878 Notes

879 See also

880 Further reading

881 Citations

882 External links

883 References

884 Notes

885 See also

886 Further reading

887 Citations

888 External links

889 References

890 Notes

891 See also

892 Further reading

893 Citations

894 External links

895 References

896 Notes

897 See also

898 Further reading

899 Citations

900 External links

901 References

902 Notes

903 See also

904 Further reading

905 Citations

906 External links

907 References

908 Notes

909 See also

910 Further reading

911 Citations

912 External links

913 References

914 Notes

915 See also

916 Further reading

917 Citations

918 External links

919 References

920 Notes

921 See also

922 Further reading

923 Citations

924 External links

925 References

926 Notes

927 See also

928 Further reading

929 Citations

930 External links

931 References

932 Notes

933 See also

934 Further reading

935 Citations

936 External links

937 References

938 Notes

939 See also

940 Further reading

941 Citations

942 External links

943 References

944 Notes

945 See also

946 Further reading

947 Citations

948 External links

949 References

950 Notes

951 See also

952 Further reading

953 Citations

954 External links

955 References

956 Notes

957 See also

958 Further reading

959 Citations

960 External links

961 References

962 Notes

963 See also

964 Further reading

965 Citations

966 External links

967 References

968 Notes

969 See also

970 Further reading

971 Citations

972 External links

973 References

974 Notes

975 See also

976 Further reading

977 Citations

978 External links

979 References

980 Notes

981 See also

982 Further reading

983 Citations

984 External links

985 References

986 Notes

987 See also

988 Further reading

989 Citations

990 External links

991 References

992 Notes

993 See also

994 Further reading

995 Citations

996 External links

997 References

998 Notes

999 See also

1000 Further reading

1001 Citations

1002 External links

1003 References

1004 Notes

1005 See also

1006 Further reading

1007 Citations

1008 External links

1009 References

1010 Notes

1011 See also

1012 Further reading

1013 Citations

1014 External links

1015 References

1016 Notes

1017 See also

1018 Further reading

1019 Citations

1020 External links

1021 References

1022 Notes

1023 See also

1024 Further reading

1025 Citations

1026 External links

1027 References

1028 Notes

1029 See also

1030 Further reading

1031 Citations

1032 External links

1033 References

1034 Notes

1035 See also

1036 Further reading

1037 Citations

1038 External links

1039 References

1040 Notes

1041 See also

1042 Further reading

1043 Citations

1044 External links

1045 References

1046 Notes

1047 See also

1048 Further reading

1049 Citations

1050 External links

1051 References

1052 Notes

1053 See also

1054 Further reading

1055 Citations

1056 External links

1057 References

1058 Notes

1059 See also

1060 Further reading

1061 Citations

1062 External links

1063 References

1064 Notes

1065 See also

1066 Further reading

1067 Citations

1068 External links

1069 References

1070 Notes

1071 See also

1072 Further reading

1073 Citations

1074 External links

1075 References

1076 Notes

1077 See also

1078 Further reading

1079 Citations

1080 External links

1081 References

1082 Notes

1083 See also

1084 Further reading

1085 Citations

1086 External links

1087 References

1088 Notes

1089 See also

1090 Further reading

1091 Citations

1092 External links

1093 References

1094 Notes

1095 See also

1096 Further reading

1097 Citations

1098 External links

1099 References

1100 Notes

1101 See also

1102 Further reading

1103 Citations

1104 External links

1105 References

1106 Notes

1107 See also

1108 Further reading

1109 Citations

1110 External links

1111 References

1112 Notes

1113 See also

1114 Further reading

1115 Citations

1116 External links

1117 References

1118 Notes

1119 See also

1120 Further reading

1121 Citations

1122 External links

1123 References

1124 Notes

1125 See also

1126 Further reading

1127 Citations

1128 External links

1129 References

1130 Notes

1131 See also

1132 Further reading

1133 Citations

1134 External links

1135 References

1136 Notes

1137 See also

1138 Further reading

1139 Citations

1140 External links

1141 References

1142 Notes

1143 See also

1144 Further reading

1145 Citations

1146 External links

1147 References

1148 Notes

1149 See also

1150 Further reading

1151 Citations

1152 External links

1153 References

1154 Notes

1155 See also

1156 Further reading

1157 Citations

1158 External links

1159 References

1160 Notes

1161 See also

1162 Further reading

1163 Citations

1164 External links

1165 References

1166 Notes

1167 See also

1168 Further reading

1169 Citations

1170 External links

1171 References

1172 Notes

1173 See also

1174 Further reading

1175 Citations

1176 External links

1177 References

1178 Notes

1179 See also

1180 Further reading

1181 Citations

1182 External links

1183 References

1184 Notes

1185 See also

1186 Further reading

1187 Citations

1188 External links

1189 References

1190 Notes

1191 See also

1192 Further reading

1193 Citations

1194 External links

1195 References

1196 Notes

1197 See also

1198 Further reading

1199 Citations

1200 External links

1201 References

1202 Notes

1203 See also

1204 Further reading

1205 Citations

1206 External links

1207 References

1208 Notes

1209 See also

1210 Further reading

1211 Citations

1212 External links

1213 References

1214 Notes

1215 See also

1216 Further reading

1217 Citations

1218 External links

1219 References

1220 Notes

1221 See also

1222 Further reading

1223 Citations

1224 External links

1225 References

1226 Notes

1227 See also

1228 Further reading

1229 Citations

1230 External links

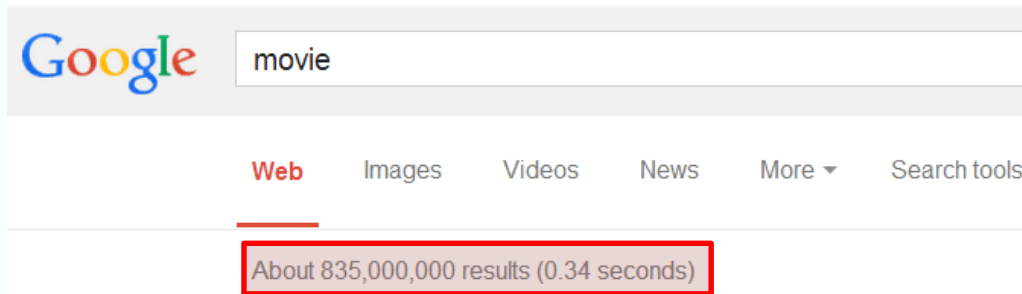
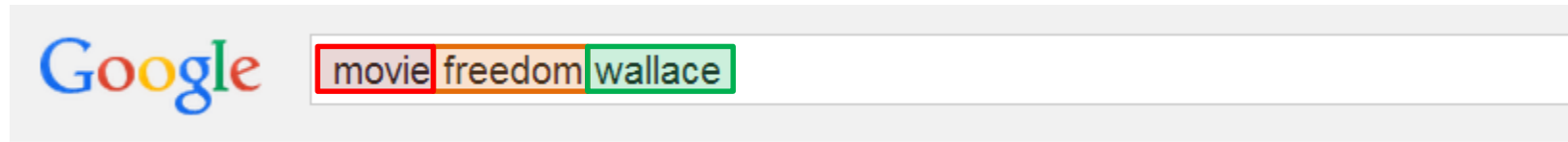
1231 References

1232 Notes

1233 See also

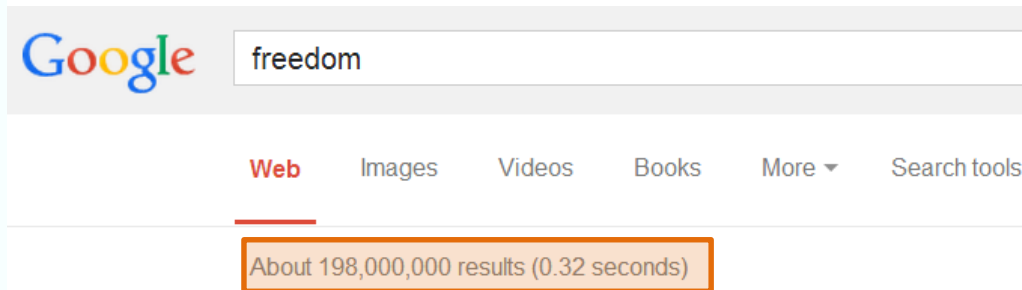
1234 Further reading

Usefulness of Words



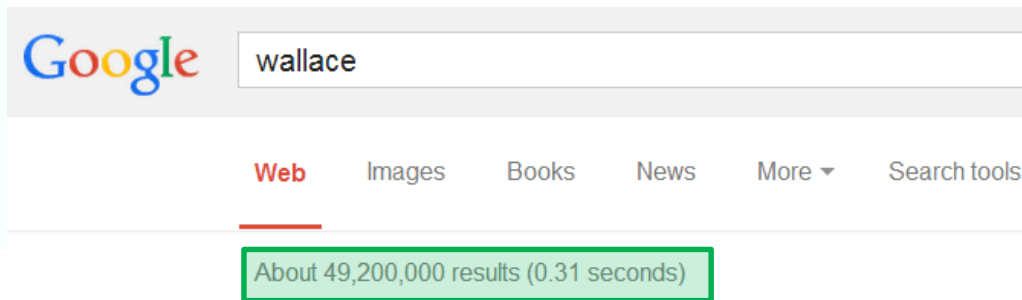
movie

- occurs very frequently



freedom

- occurs frequently



wallace

- occurs occassionally

Estimating Relevance

- Rare words more important than common words
 - wallace (49M) more important than freedom (198M)
more important than movie (835M)
- Words occurring more frequently in a document indicate higher relevance
 - wallace (88) more matches than movie (16) more matches than freedom (7)

Relevance Measure: TF-IDF

- TF: Term Frequency
 - Measures occurrences of a term in a document
 - $\text{tf}(t, d)$... various options

- Raw count of occurrences

$$\text{tf}(t, d) = \text{count}(t, d)$$

- Logarithmically scaled

$$\text{tf}(t, d) = \log(\text{count}(t, d) + 1)$$

- Normalised by document length

$$\text{tf}(t, d) = \frac{\text{count}(t, d)}{\sum_{t' \in d} \text{count}(t', d)}$$

$$\text{tf}(t, d) = \frac{\text{count}(t, d)}{\max_{t' \in d} \text{count}(t', d)}$$

- A combination / something else 😊

Relevance Measure: TF-IDF

- **IDF**: Inverse Document Frequency
 - How common a term is across **all** documents
 - $\text{idf}(t, D)$...
 - Logarithmically scaled document occurrences

$$\text{idf}(t, D) = \log\left(\frac{|D|+1}{|\{d \in D : t \in d\}|+1}\right)$$

- Note: The more rare, the larger the value

Relevance Measure: TF-IDF

- TF-IDF: Combine Term Frequency and Inverse Document Frequency:

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t, D)$$

- Score for a query
 - Let query $q = (t_1, \dots, t_n)$
 - Score for a query: $\text{score}(q, d) = \sum_{t \in q} \text{tf-idf}(t, d)$
- (There are other possibilities)

Relevance Measure: TF-IDF



Term Frequency

$$\text{tf}(t, d) = \text{count}(t, d)$$

Inverse Document Frequency

$$\text{idf}(t, D) = \log_2 \left(\frac{|D|}{|\{d \in D : t \in d\}| + 1} \right)$$

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t, D)$$

t	$\text{tf}(t, d)$
movie	16
freedom	7
wallace	43

Relevance Measure: TF-IDF



Term Frequency

$$\text{tf}(t, d) = \text{count}(t, d)$$

Inverse Document Frequency

$$\text{idf}(t, D) = \log_2 \left(\frac{|D|}{|\{d \in D : t \in d\}| + 1} \right)$$

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t, D)$$

t	$\text{tf}(t, d)$	$ \{d \in D : t \in d\} $
movie	16	835,000,000
freedom	7	198,000,000
wallace	43	49,200,000

Relevance Measure: TF-IDF



Term Frequency

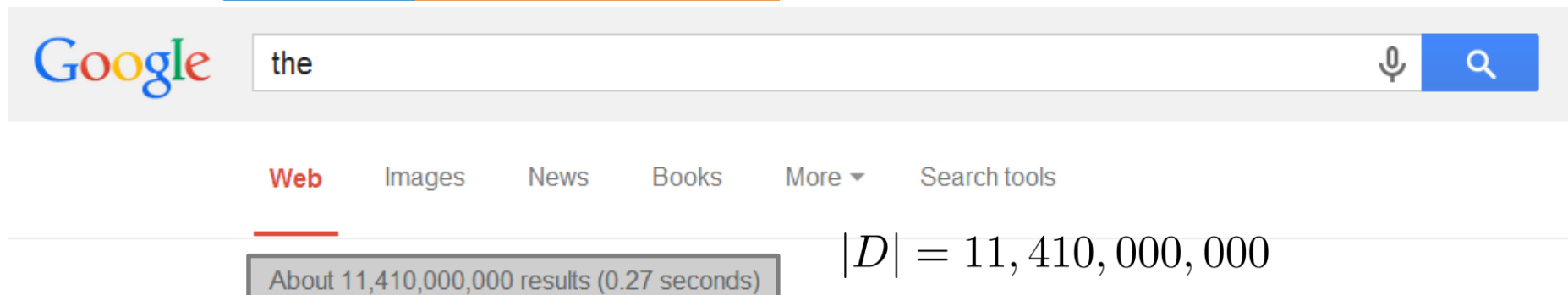
$$\text{tf}(t, d) = \text{count}(t, d)$$

Inverse Document Frequency

$$\text{idf}(t, D) = \log_2 \left(\frac{|D|}{|\{d \in D : t \in d\}| + 1} \right)$$

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t, D)$$

t	$\text{tf}(t, d)$	$ \{d \in D : t \in d\} $	$\frac{ D }{ \{d \in D : t \in d\} + 1}$
movie	16	835,000,000	
freedom	7	198,000,000	
wallace	43	49,200,000	



Relevance Measure: TF-IDF



Term Frequency

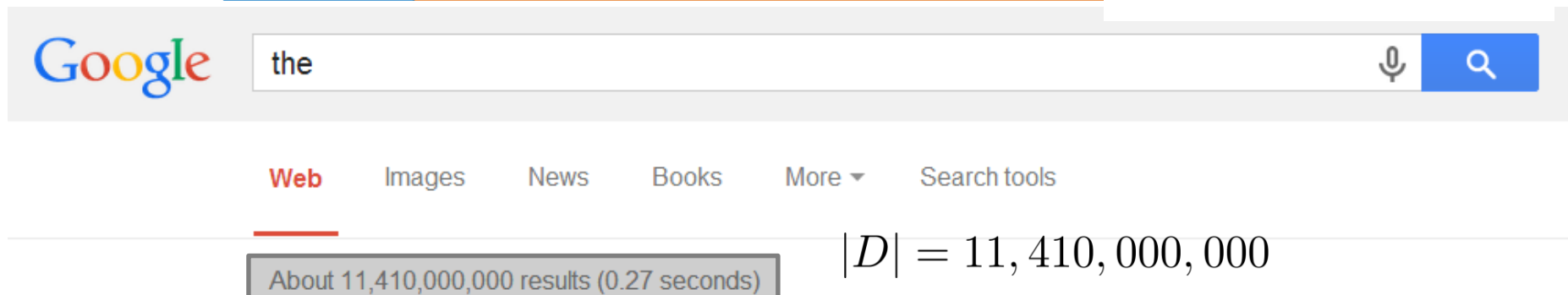
$$\text{tf}(t, d) = \text{count}(t, d)$$

Inverse Document Frequency

$$\text{idf}(t, D) = \log_2 \left(\frac{|D|}{|\{d \in D : t \in d\}| + 1} \right)$$

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t, D)$$

t	$\text{tf}(t, d)$	$ \{d \in D : t \in d\} $	$\frac{ D }{ \{d \in D : t \in d\} + 1}$
movie	16	835,000,000	13.66
freedom	7	198,000,000	57.62
wallace	43	49,200,000	231.91



$$|D| = 11,410,000,000$$

Relevance Measure: TF-IDF



Term Frequency

$$\text{tf}(t, d) = \text{count}(t, d)$$

Inverse Document Frequency

$$\text{idf}(t, D) = \log_2 \left(\frac{|D|}{|\{d \in D : t \in d\}| + 1} \right)$$

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t, D)$$

t	$\text{tf}(t, d)$	$ \{d \in D : t \in d\} $	$\frac{ D }{ \{d \in D : t \in d\} + 1}$	$\text{idf}(t, D)$
movie	16	835,000,000	13.66	3.77
freedom	7	198,000,000	57.62	5.84
wallace	43	49,200,000	231.91	7.85

Relevance Measure: TF-IDF



Term Frequency

$$\text{tf}(t, d) = \text{count}(t, d)$$

Inverse Document Frequency

$$\text{idf}(t, D) = \log_2 \left(\frac{|D|}{|\{d \in D : t \in d\}| + 1} \right)$$

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t, D)$$

t	$\text{tf}(t, d)$	$ \{d \in D : t \in d\} $	$\frac{ D }{ \{d \in D : t \in d\} + 1}$	$\text{idf}(t, D)$	$\text{tf-idf}(t, d)$
movie	16	835,000,000	13.66	3.77	60.36
freedom	7	198,000,000	57.62	5.84	40.94
wallace	43	49,200,000	231.91	7.85	337.87

Relevance Measure: TF-IDF



Term Frequency

$$\text{tf}(t, d) = \text{count}(t, d)$$

Inverse Document Frequency

$$\text{idf}(t, D) = \log_2 \left(\frac{|D|}{|\{d \in D : t \in d\}| + 1} \right)$$

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t, D)$$

t	$\text{tf}(t, d)$	$ \{d \in D : t \in d\} $	$\frac{ D }{ \{d \in D : t \in d\} + 1}$	$\text{idf}(t, D)$	$\text{tf-idf}(t, d)$
movie	16	835,000,000	13.66	3.77	60.36
freedom	7	198,000,000	57.62	5.84	40.94
wallace	43	49,200,000	231.91	7.85	337.87

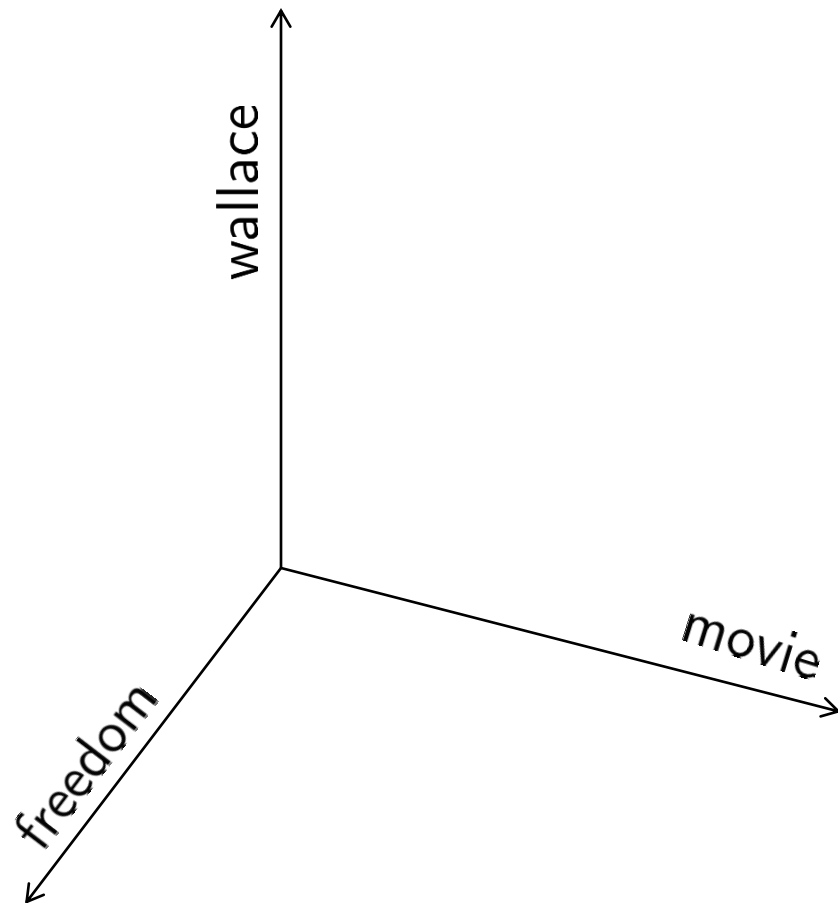
$$\text{score}(q, d) = \sum_{t \in q} \text{tf-idf}(t, d)$$

$$\text{score}((\text{movie, freedom, wallace}), \text{http://en.wikipedia.org/Braveheart}) \approx 439.17$$

Vector Space Model (a mention)

t	$\text{tf}(t, d)$
movie	16
freedom	7
wallace	43

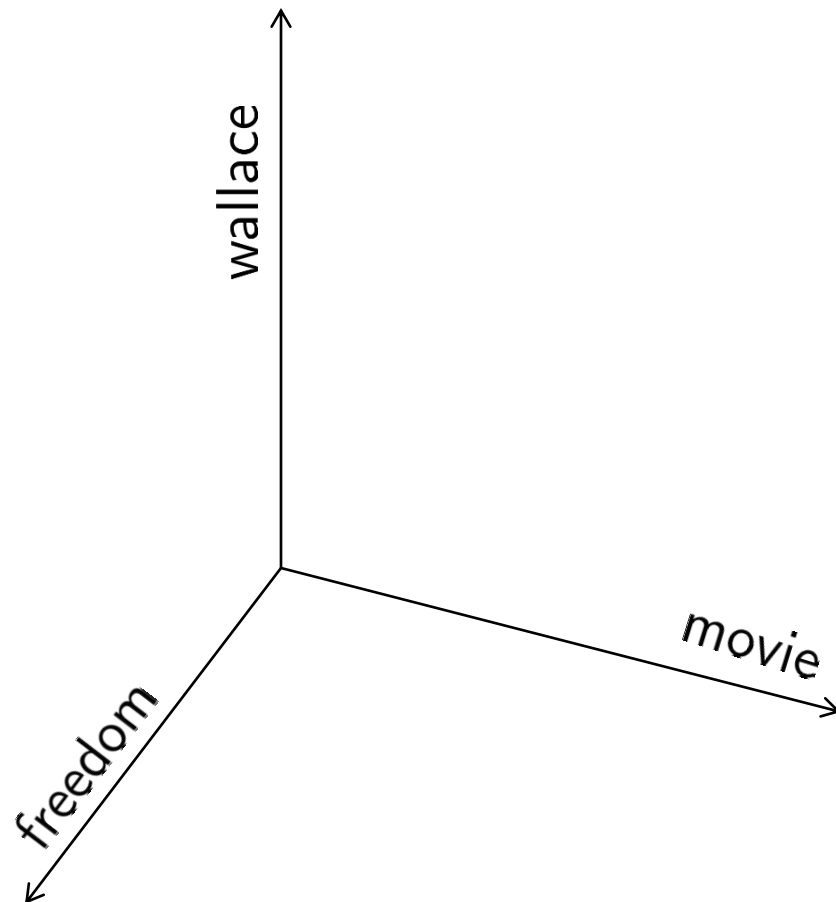
$$l = \sqrt{\sum_{t \in q} \text{tf}(t, d)^2}$$



Vector Space Model (a mention)

t	$\text{tf}(t, d)$	$\text{tf}(t, d)^2$
movie	16	256
freedom	7	49
wallace	43	1,894

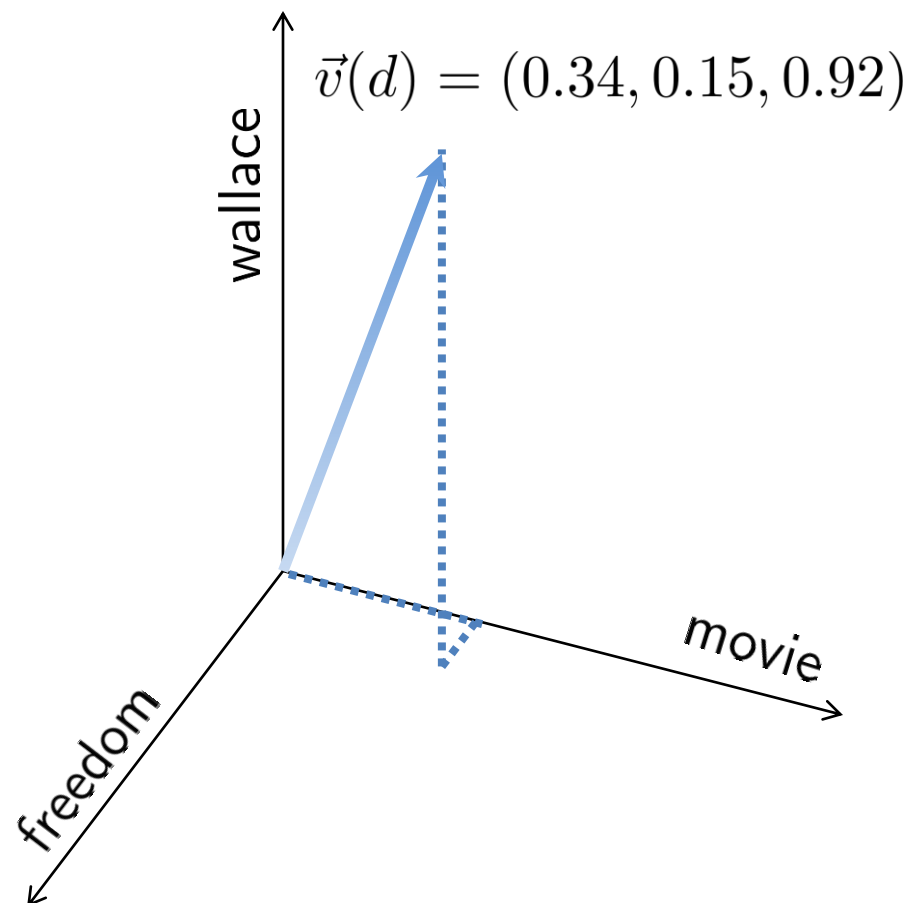
$$l = \sqrt{\sum_{t \in q} \text{tf}(t, d)^2}$$



Vector Space Model (a mention)

t	$\text{tf}(t, d)$	$\text{tf}(t, d)^2$	$\frac{\text{tf}(t, d)}{l}$
movie	16	256	0.34
freedom	7	49	0.15
wallace	43	1,894	0.92

$$l = \sqrt{\sum_{t \in q} \text{tf}(t, d)^2}$$



Dividing by l normalises the length of vector to 1

Vector Space Model (a mention)

- Cosine Similarity

$$\text{sim}(d, d') = \vec{v}(d) \cdot \vec{v}(d')$$

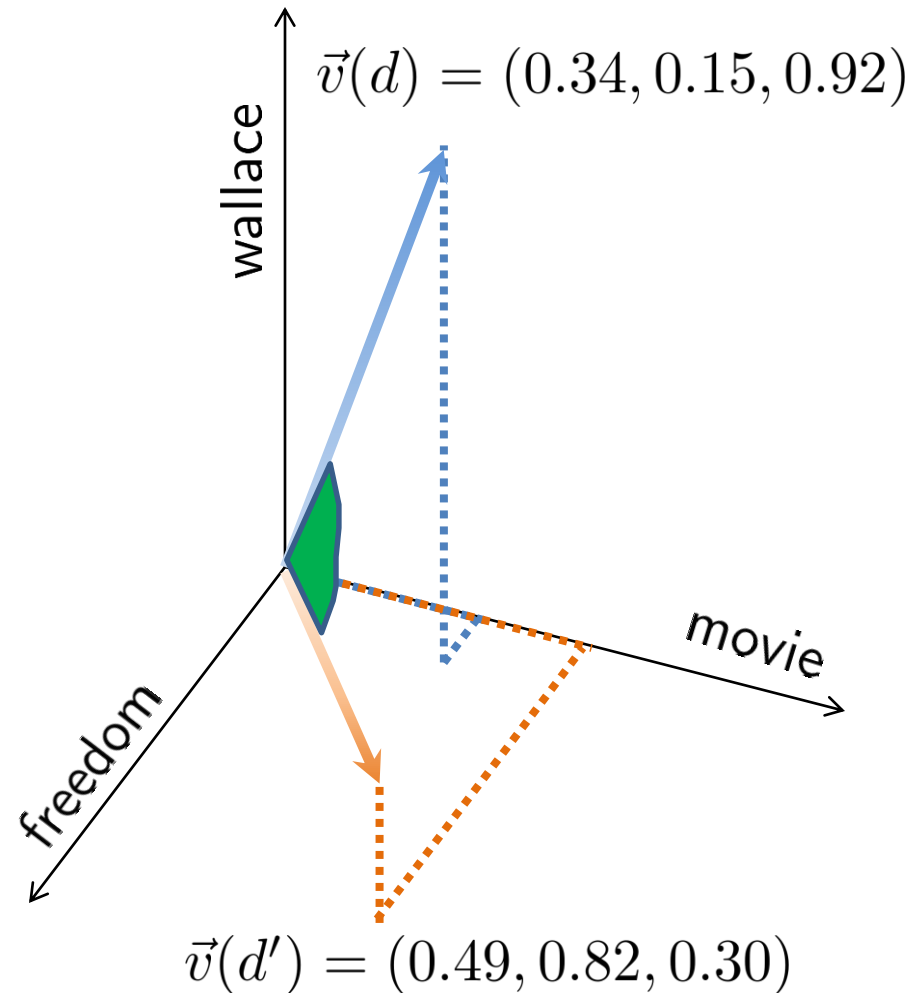
t	$\vec{v}(d)$	$\vec{v}(d')$	\times
movie	0.34	0.49	0.17
freedom	0.15	0.82	0.12
wallace	0.93	0.30	0.28

$$\text{sim}(d, d') \approx 0.57$$

- Note:

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos(\angle(\mathbf{a}, \mathbf{b}))$$

$$|\vec{v}(d)| = |\vec{v}(d')| = 1$$



Hence the similarity is the cosine of the **angle** between the vectors

Relevance Measure: TF-IDF

- TF-IDF: Combine Term Frequency and Inverse Document Frequency:

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t, D)$$

- Score for a query
 - Let query $q = (t_1, \dots, t_n)$
 - Score for a query: $\text{score}(q, d) = \sum_{t \in q} \text{tf-idf}(t, d)$(There are other possibilities)

... we could also use cosine similarity between query and document using TF-IDF weights

Two Sides to Ranking: Relevance



The image shows a Google search interface with the query 'obama' in the search bar. The search results are displayed under the 'Web' tab, showing approximately 16,700,000 results in 0.23 seconds. The top results are for 'Broccoli - Wikipedia, the free encyclopedia' and 'Broccoli - The World's Healthiest Foods'. A third result, 'News for broccoli', is partially visible. A large red 'X' is drawn over the search results. On the left side of the image, there is a photo of Barack Obama, and on the right side, there is a head of broccoli.

Google

obama

Web Images News Videos More ▾ Search tools

About 16,700,000 results (0.23 seconds)

Broccoli - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Broccoli** ▾
Broccoli is an edible green plant in the cabbage family, whose large flowering head is used as a vegetable. The word **broccoli** comes from the Italian plural of ...
Cauliflower - Romanesco broccoli - Broccoli (disambiguation) - Brocolini

Broccoli - The World's Healthiest Foods
www.whfoods.com/genpage.php?tname=foodspice&dbid=9 ▾
Broccoli can provide you with some special cholesterol-lowering benefits if you will cook it by steaming. The fiber-related components in **broccoli** do a better job ...

News for broccoli

Mistakes We All Make With Spaghetti, Steak And E
Huffington Post - 2 days ago
But in her new book Brassicas: Cooking the World's Healthiest Vegetables, she says plunking **broccoli**, cauliflower or Brussels sprouts into ...

Field-Based Boosting

- Not all text is equal: titles, headers, etc.

```
<!DOCTYPE html>
<html lang="en" dir="ltr" class="client-nojs">
<head>
<meta charset="UTF-8" />
<title>Barack Obama - Wikipedia, the free encyclopedia</title>
```



The screenshot displays the Wikipedia article for Barack Obama. The title "Barack Obama" is highlighted with an orange box. The introductory text states: "Barack Hussein Obama II (ⁱˈbəˈrɑːk huːˈseɪn oʊˈbɑːmə; born August 4, 1961) is the 44th and current President of the United States, and the first African American to hold the office. Born in Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he served as president of the Harvard Law Review. He was a community organizer in Chicago before earning his law degree. He worked as a civil rights attorney and taught constitutional law at the University of Chicago Law School.

Barack Obama

From Wikipedia, the free encyclopedia

"Obama" redirects here. For other uses, see [Obama \(disambiguation\)](#).

This article is about the 44th president of the United States. For his father, see [Barack Obama, Sr.](#)

Barack Hussein Obama II (ⁱˈbəˈrɑːk huːˈseɪn oʊˈbɑːmə; born August 4, 1961) is the 44th and current President of the United States, and the first African American to hold the office. Born in Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he served as president of the Harvard Law Review. He was a community organizer in Chicago before earning his law degree. He worked as a civil rights attorney and taught constitutional law at the University of Chicago Law School.



Anchor Text

- See how the Web views/tags a page

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html>
<head>
  <title>What I watched last night ...</title>
</head>
<body>
<p>Last night I was pretty bored so I made some popcorn and watched
<a href="http://en.wikipedia.org/Braveheart">a movie about William Wallace called Braveheart</a>.
Set in Scotland it has lots of blood and gore.
</p>
</body>
</html>
```

Anchor Text

- See how the Web views/tags a page


```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/
<html>
<head>
  <title>What I watched
</head>
<body>
<p>Last night I was pret
<a href="http://en.wiki
Set in Scotland it has
</p>
</body>
</html>
```

Google

da da da dum symphony

Web Videos News Shopping Images More Search tools

About 107,000 results (0.36 seconds)



Beethoven - Symphony No. 5 in C Minor (1) - YouTube

www.youtube.com/watch?v=W2qW6fOtAMY

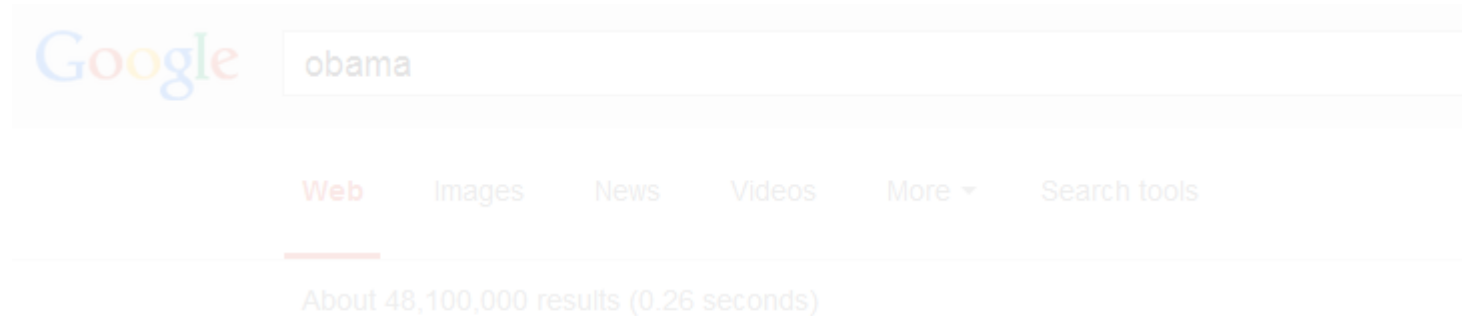
Lucene uses relevance scoring



```
Tasks Console
SearchWikiIndex [Java Application] C:\Program Files\Java\jre1.8.0_65\bin\javaw.exe (03-05-2017 12:45:22 a. m.)
Opening directory at lucene
Enter a keyword search phrase:
obama
Running query: obama
Parsed query: TITLE:obam^5.0 ABSTRACT:obam
Matching documents: 255
Showing top 10 results
1 http://es.wikipedia.org/wiki/Obama_Republican Obama Republican
2 http://es.wikipedia.org/wiki/Obama_(Fukui) Obama (Fukui)
3 http://es.wikipedia.org/wiki/Republicanos_por_Obama Republicanos por Obama
4 http://es.wikipedia.org/wiki/Engonga_Obame Engonga Obame
5 http://es.wikipedia.org/wiki/Barack_Obama Barack Obama
6 http://es.wikipedia.org/wiki/Michelle_Obama Michelle Obama
7 http://es.wikipedia.org/wiki/Cartel_%22Hope%22_de_Obama Cartel "Hope" de Obama
8 http://es.wikipedia.org/wiki/Transición_presidencial_de_Barack_Obama Transición presidencial de Barack Obama
9 http://es.wikipedia.org/wiki/Por_qué_Obama_ganará_en_2008_y_en_2012 Por qué Obama ganará en 2008 y en 2012
10 http://es.wikipedia.org/wiki/Ricardo_Mangue_Obama_Nfubea Ricardo Mangue Obama Nfubea
```

RANKING:
IMPORTANCE

Two Sides to Ranking: Importance

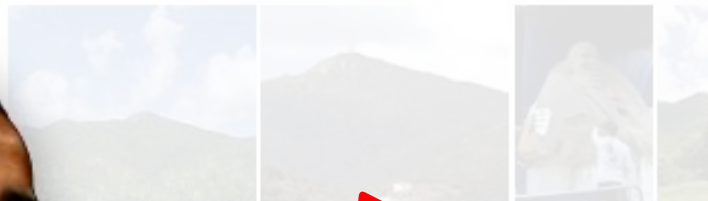


How could we determine that Barack Obama is more important than Mount Obama as a search result for "obama" on the Web?



Images for mount obama

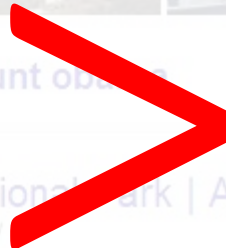
Report images



More images for mount obama

Mount Obama National Park | Antigua and Barbuda
antiguamountobama.com/

Jun 16, 2011 - As the Mount Obama Committee continues its work in the Area, the committee organized a site visit to the O



Link Analysis

Which will have more links from other pages?
The Wikipedia article for Mount Obama?
The Wikipedia article for Barack Obama?



Link Analysis

- Consider links as votes of confidence in a page
- A hyperlink is the open Web's version of ...



(... even if the page is linked in a negative way.)

Link Analysis

So if we just count links to a page we can determine its importance and we are done?



Link Spamming



semanticweb.com™

The Voice of Semantic Technology Business
Big Data, Linked Data, Smart Data

Home

Events

Media

Industry Verticals

Answers

Jobs

Questions

Tags

Users

Badges

[deleted] Kala Jadu Specialist +91961

Qu



black magic specialist baba ji call now +919610897260



<http://www.blackmagicspecialist.net.in>



java

edit | close | undelete | more ▼

[Claritin](#) [Clomid](#) [Combivent](#) [Confido](#) [Copegus](#) [Cordarone](#) [Coreg](#) [Coumadin](#) [Cozaar](#) [Crestor](#)
[Cyklokapron](#) [Cymbalta](#) [Cystone](#) [Cytotec](#) [Danazol](#) [Deltasone](#) [Depakote](#) [Desyrel](#) [Detrol](#) [Diabecor](#)
[Diakof](#) [Diarex](#) [Didronel](#) [Differin](#) [Dilantin](#) [Diovan](#) [Dostinex](#) [Elavil](#) [Elimite](#) [Emsam](#) [Endep](#) [Eurax](#)
[Evicare](#) [Evista](#) [Exelon](#) [Famvir](#) [Feldene](#) [Femara](#) [Femcare](#) [Flomax](#) [Flonase](#) [Flovent](#) [Fosamax](#) [Gasex](#)
[Geodon](#) [Geriforte](#) [Herbolax](#) [High Love](#) [Himcoid](#) [Himcolin](#) [Himcospaz](#) [Himplasia](#) [Hoodia](#) [Hytrin](#)
[Hyzaar](#) [Imdur](#) [Imitrex](#) [Inderal](#) [Ismo](#) [Isoptin](#) [Isordil](#) [Kamagra](#) [Karela](#) [Keftab](#) [Koflet](#) [Kytril](#) [Lamictal](#)
[Lamisil](#) [Lanoxin](#) [Lariam](#) [Lasix](#) [Lasuna](#) [Leukeran](#) [Levaquin](#) [Levlen](#) [Levothroid](#) [Lincocin](#) [Lioresal](#)
[Lisinopril](#) [Liv.52](#) [Lopid](#) [Lopressor](#) [Loprox](#) [Lotensin](#) [Lotrisone](#) [Loxitane](#) [Lozol](#) [Lukol](#) [Lynoral](#)
[Maxaquin](#) [Menosan](#) [Mentat](#) [Mentax](#) [Mevacor](#) [Mexitil](#) [Miacalcin](#) [Micardis](#) [Mobic](#) [Monoket](#) [Motrin](#)
[Myambutol](#) [Myclex-G](#) [Mysoline](#) [Naprosyn](#) [Neurontin](#) [Nicotinell](#) [Nimotop](#) [Nirdosh](#) [Nizoral](#)
[Nolvadex](#) [Nonoxinol](#) [Noroxin](#) [Omnicef](#) [Ophthalmicare](#) [Oxytrol](#) [Pamelor](#) [Parlodel](#) [Paxil](#) [Penisole](#)
[Phentermine](#) [Pilex](#) [Plan B](#) [Plavix](#) [Plendil](#) [Pletal](#) [Prandin](#) [Pravachol](#) [Prednisone](#) [Pregamin](#) [Prevacid](#)
[Prilosec](#) [Prinivil](#) [Procardia](#) [Prograf](#) [Prometrium](#) [Propecia](#) [Proscar](#) [Protonix](#) [Proventil](#) [Prozac](#) [Purin](#)
[Purinethol](#) [Quibron-T](#) [Relafen](#) [Renalka](#) [Reosto](#) [Requip](#) [Retin-A](#) [Revvia](#) [Rhinocort](#) [Rimonabant](#)
[Risperdal](#) [Rocaltrol](#) [Rogaine](#) [Rumalaya](#) [Sarafem](#) [Septilin](#) [Serevent](#) [Serophene](#) [Seroquel](#) [Shallaki](#)
[Shoot](#) [Sinequan](#) [Singulair](#) [Snoroff](#) [Sorbitrate](#) [Speman](#) [Starlix](#) [Stretchit](#) [Stromectol](#) [Styplon](#)
[Sumycin](#) [Superman](#) [Sustiva](#) [Synthroid](#) [Tenormin](#) [Topamax](#) [Trandate](#) [Tricor](#) [Trimox](#) [Triphala](#) [Tulasi](#)
[Urispas](#) [V-Gel](#) [Vantin](#) [Vasodilan](#) [Vasotec](#) [Ventolin](#) [Viramune](#) [Vytorin](#) [Xeloda](#) [Xenacore](#) [Zanaflex](#)
[Zantac](#) [Zebeta](#) [Zelnorm](#) [Zerit](#) [Yerba Diet](#) [Wellbutrin SR](#) [Women Attracting Pheromones](#) [Women's](#)
[Intimacy Enhancer](#) [Women's Intimacy Enhancer Cream](#) [Virility Gum](#) [Vitamin A & D](#) [Viagra +](#)
[Cialis](#) [Viagra + Cialis](#) [Levitra](#) [Viagra Jelly](#) [Viagra Soft](#) [Cialis Soft](#) [Viagra Soft Tabs](#) [Ultimate](#)
[Male Enhancer](#) [Toprol XL](#) [Touch-Up Kit](#) [Tentex Royal](#) [Tentex Forte](#) [Tiberius Erectus](#) [Zero](#)
[Nicotine 2](#) [Complete Professional Whitening Kits 2](#) [Sets Of Moldable Mouth Trays 36](#) [Beauty](#)
[Acne-n-Pimple Cream](#) [ActoPlus Met](#) [Superloss Multi](#) [Sleep Well \(Herbal XANAX\)](#) [Shuddha](#)
[Guggulu](#) [Rythmol SR](#) [Rumalaya Forte](#) [Pulmicort Inhaler](#) [Professional Plasma Tooth Whitening Kit](#)
[Premium Diet Patch](#) [Penis Growth Oil](#) [Penis Growth Pack](#) [Penis Growth Patch](#) [Penis Growth Pills](#)
[Orgasm Enhancer](#) [Norpace CR](#) [Mental Booster](#) [Men Attracting Pheromones](#) [Menopause Gum](#)
[Male Enhancement Oil](#) [Male Enhancement Patch](#) [Male Enhancement Pills](#) [Male Sexual Tonic](#)
[InnoPran XL](#) [Hoodia Weght Loss Gum](#) [Hoodia Weight Loss Patch](#) [Human Growth Hormone](#)
[Agent](#) [Glucotrol XL](#) [Green Tea](#) [Grifulvin V](#) [Gyne-Lotrimin](#) [Hair Loss Cream](#) [Herbal Maxx](#) [Herbal](#)
[Phentermine](#) [Flagyl ER](#) [Female Sexual Tonic](#) [Female Viagra](#) [Epivir-HBV](#) [Diet Maxx](#) [Deluxe](#)
[Handheld Plasma Whitening Tool](#) [Deluxe Whitening System With Plasma Lamp](#) [Coral Calcium](#)
[Cialis Jelly](#) [Cialis Soft Tabs](#) [Calcium Carbonate](#) [Bust Enhancer](#) [Beconase AQ](#) [Anatrim Diet Pills](#)
[Advair Diskus](#) [Advanced Gain Pro](#) [Breast Augmentation](#) [Breast Enhancement](#) [Breast Enhancement](#)
[Gel](#) [Breast Enhancement Gum](#) [Breast Intense Buy](#) [Trazodone Buy](#) [Celebrex Buy](#) [Alprazolam Buy](#)
[Tramadol Buy](#) [Fioricet Buy](#) [Soma Buy](#) [Cialis Buy](#) [Carisoprodol Buy](#) [Levitra Buy](#) [Ultram Buy](#)
[Ambien Buy](#) [Viagra Buy](#) [Xanax Buy](#) [Phentermine Buy](#) [Valium Buy](#) [Diazepam](#) [Generic Celebrex](#)
[Generic Alprazolam](#) [Generic Tramadol](#) [Generic Fioricet](#) [Generic Soma](#) [Generic Cialis](#) [Generic](#)
[Carisoprodol](#) [Generic Levitra](#) [Generic Ultram](#) [Generic Ambien](#)

Link Importance

So which should count for more?

A link from http://en.wikipedia.org/wiki/Barack_Obama?

Or a link from <http://freev1agra.com/shop.html>?



Link Importance

Maybe we could consider links from some domains as having more “vote”?



A screenshot of the Digg website interface. The top navigation bar includes the Digg logo, "My News", "Top News", and "Upcoming", along with a search bar and a "GO" button. Below this is a category bar with links for "All Topics", "Business", "Entertainment", "Gaming", "Lifestyle" (which is highlighted), "Offbeat", "Politics", "Science", and "Sports". The main content area displays a link from "rxpills.host-sc.com" with a Digg score of 385. The link title is "RxPILLS - Best discounts for all Pharmacy! Check out or men's health category!". The link text is a long, repetitive string of keywords related to Viagra, such as "viagra toll free number", "viagra tonytigeraz", "viagra top ten", "viagra toronto", "viagra torrent", "viagra total knee", "viagra toung", "viagra travel", "viagra treat childhood pulmonary hypertension", "viagra treatment", "viagra treatment for crohn's", "viagra treatment for pe", "viagra treatment for women", "viagra treatment hape", "viagra treatment impotence", "viagra treats children s lethal hypertension", "viagra trh pharmacy", "viagra trial", "viagra trial coupon", "viagra trial pack", "viagra trial pack canada", "viagra trial packs", "viagra trial pak", "viagra trial sample", "viagra trials for shrinking cancerous tumors", "viagra triangle", "viagra triangle chicago", "viagra triangle chicago illinois", "viagra triangle cleveland", "viagra triangle cleveland ohio", "viagra tricks", "viagra trip in thailand", "viagra trivia", "viagra tune", "viagra tupperware", "viagra tupperware pool", "viagra tutorial", "viagra tv ad", "viagra tv ad 1999", "viagra tv advertising", "viagra tv commercial", "viagra tv commercial girl", "viagra tv commercial viva las vegas", "viagra tv commercials", "viagra tv girl", "viagra type drugs", "viagra type medications", "viagra type products", "viagra types", "viagra u", "viagra u s pharmacies", "viagra uden receipt", "viagra uk", "viagra uk 32", "viagra uk buy online", "viagra uk cheap purchase", "viagra uk cost pill", "viagra uk delivery", "viagra uk forum", "viagra uk kamagra", "viagra uk news", "viagra uk online a href", "viagra uk online", "viagra uk viagra", "viagra uk purchase", "viagra ghuk retail price", "viagra ukasdasd retffs". The link is dated "2 days ago". To the right of the link is a small thumbnail image showing a couple.

PageRank



PageRank

- Not just a count of inlinks
 - A link from a more important page is more important
 - A link from a page with fewer links is more important
 - ∴ A page with lots of inlinks from important pages (which have few outlinks) is more important

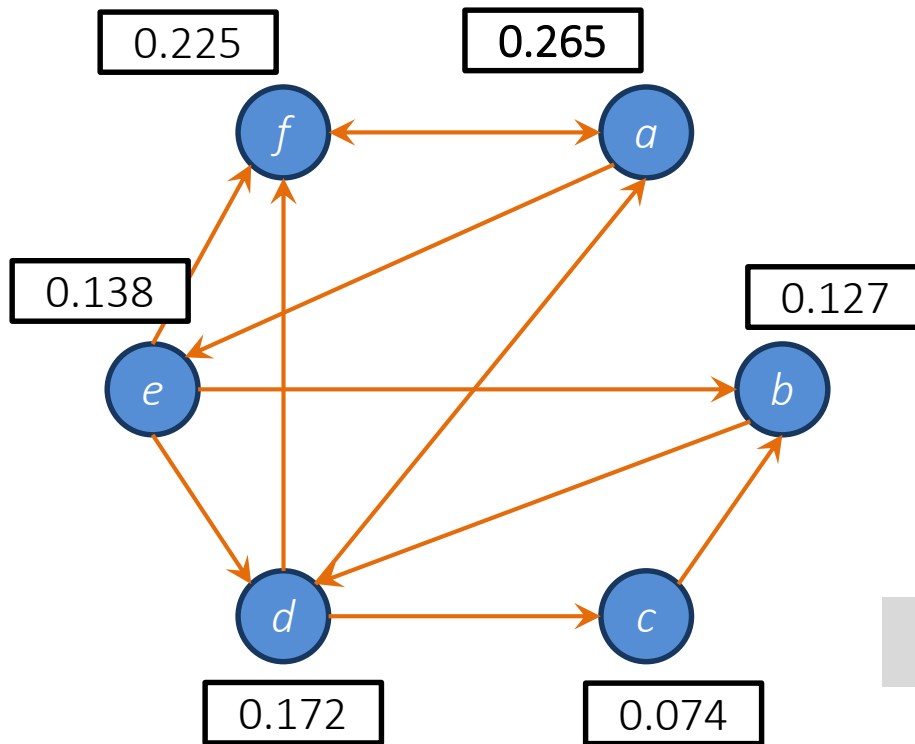
PageRank is Recursive

- Not just a count of inlinks
 - A link from a more important page is more important
 - A link from a page with lots of inlinks is less important
 - ∴ A page with lots of inlinks and few outgoing links is more important



PageRank Model

- The Web: a directed graph



$$G = [V, E]$$

Vertices
(pages)

Edges
(links)

Which vertex is most important?



$$V = \{a, b, c, d, e, f\}$$

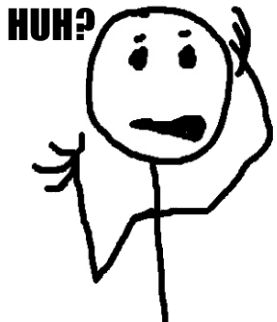
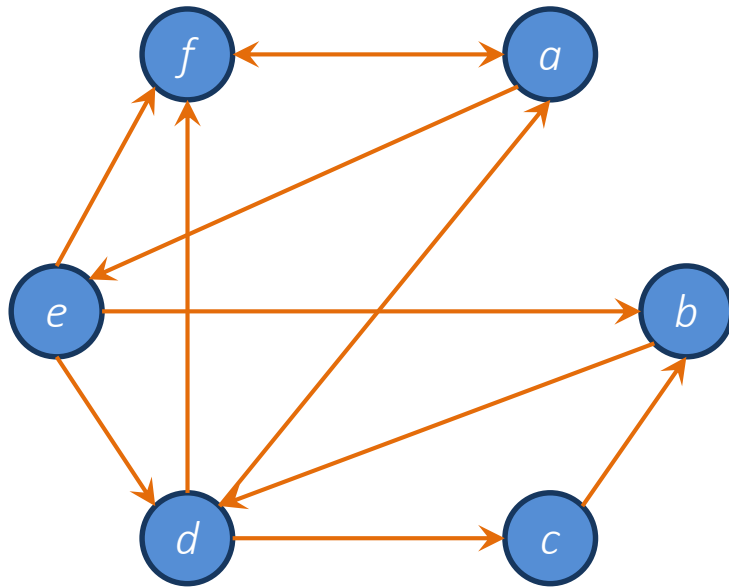
$$E = \{(a, e), (a, f), (b, d), (c, b), (d, a), (d, c), (d, f), (e, b), (e, d), (e, f), (f, a)\}$$

PageRank Model

- The Web: a directed graph

$$G = \boxed{V}, \boxed{E}$$

Vertices (pages)	Edges (links)
---------------------	------------------



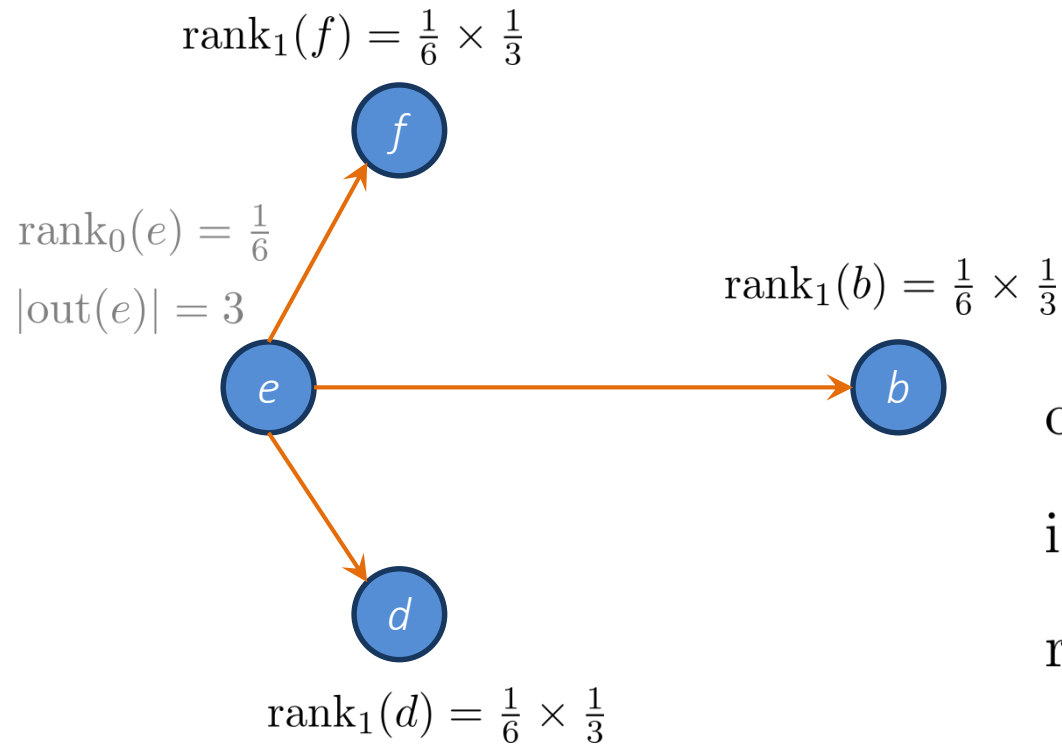
$$\text{out}(v) \doteq \{v' \in V \mid (v, v') \in E\}$$

$$\text{in}(v) \doteq \{v' \in V \mid (v', v) \in E\}$$

$$\text{rank}_0(v) \doteq \frac{1}{|V|}$$

$$\text{rank}_i(v) \doteq \sum_{v' \in \text{in}(v)} \frac{\text{rank}_{i-1}(v')}{|\text{out}(v')|}$$

PageRank Model



$$G = [V, E]$$

Vertices (pages)	Edges (links)
---------------------	------------------

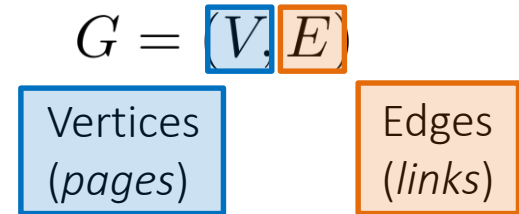
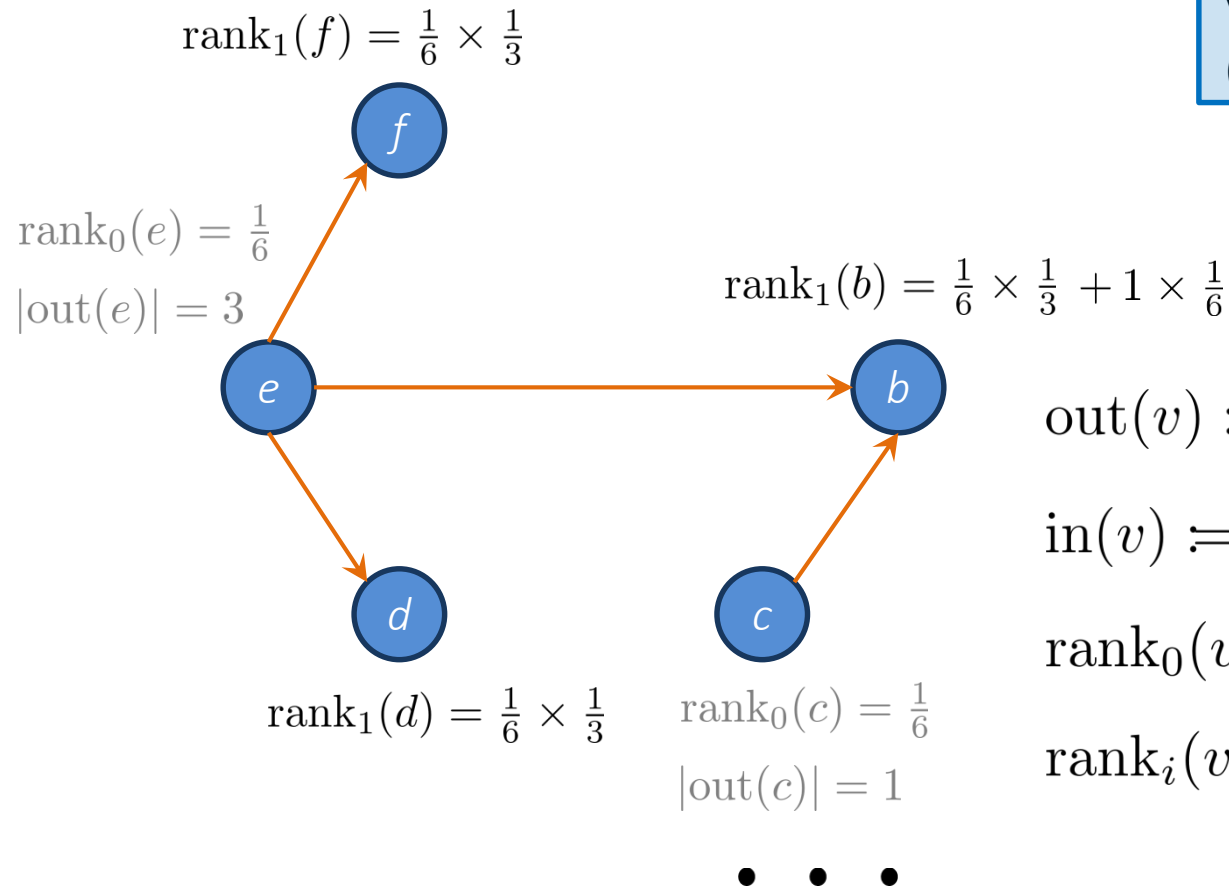
$$\text{out}(v) \doteq \{v' \in V \mid (v, v') \in E\}$$

$$\text{in}(v) \doteq \{v' \in V \mid (v', v) \in E\}$$

$$\text{rank}_0(v) \doteq \frac{1}{|V|}$$

$$\text{rank}_i(v) \doteq \sum_{v' \in \text{in}(v)} \frac{\text{rank}_{i-1}(v')}{|\text{out}(v')|}$$

PageRank Model



$$\text{out}(v) \doteq \{v' \in V \mid (v, v') \in E\}$$

$$\text{in}(v) \doteq \{v' \in V \mid (v', v) \in E\}$$

$$\text{rank}_0(v) \doteq \frac{1}{|V|}$$

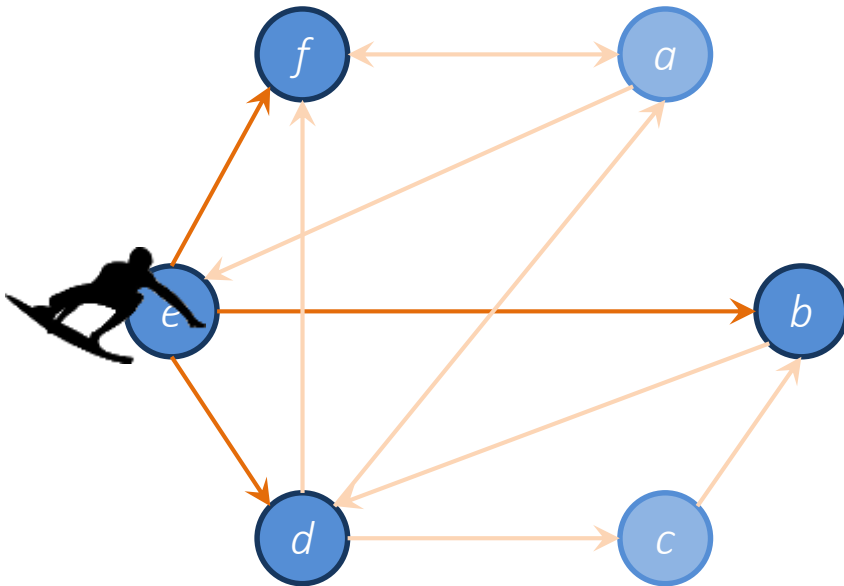
$$\text{rank}_i(v) \doteq \sum_{v' \in \text{in}(v)} \frac{\text{rank}_{i-1}(v')}{|\text{out}(v')|}$$

PageRank: Random Surfer Model



= someone surfing the web,
clicking links randomly

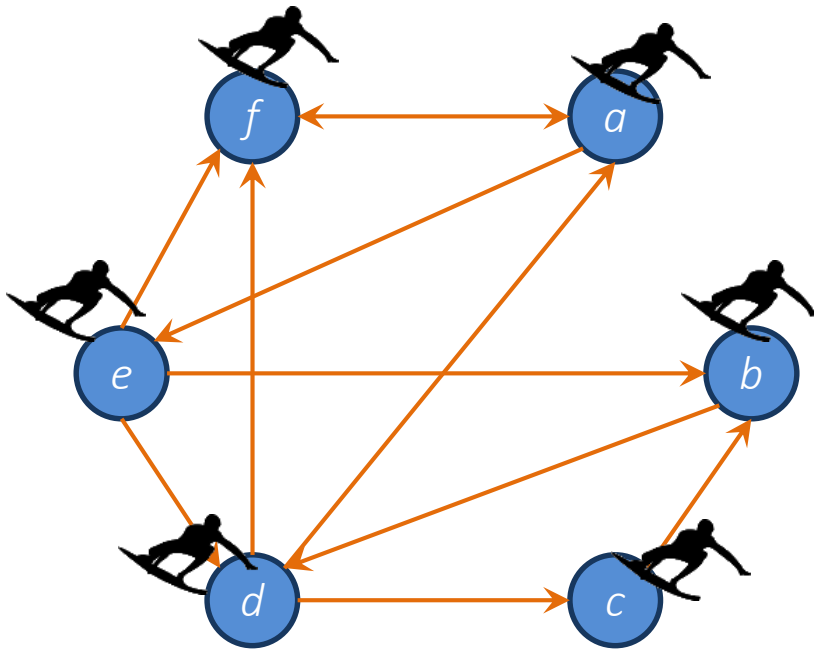
- What is the probability of being at page x after n hops?



PageRank: Random Surfer Model



= someone surfing the web,
clicking links randomly

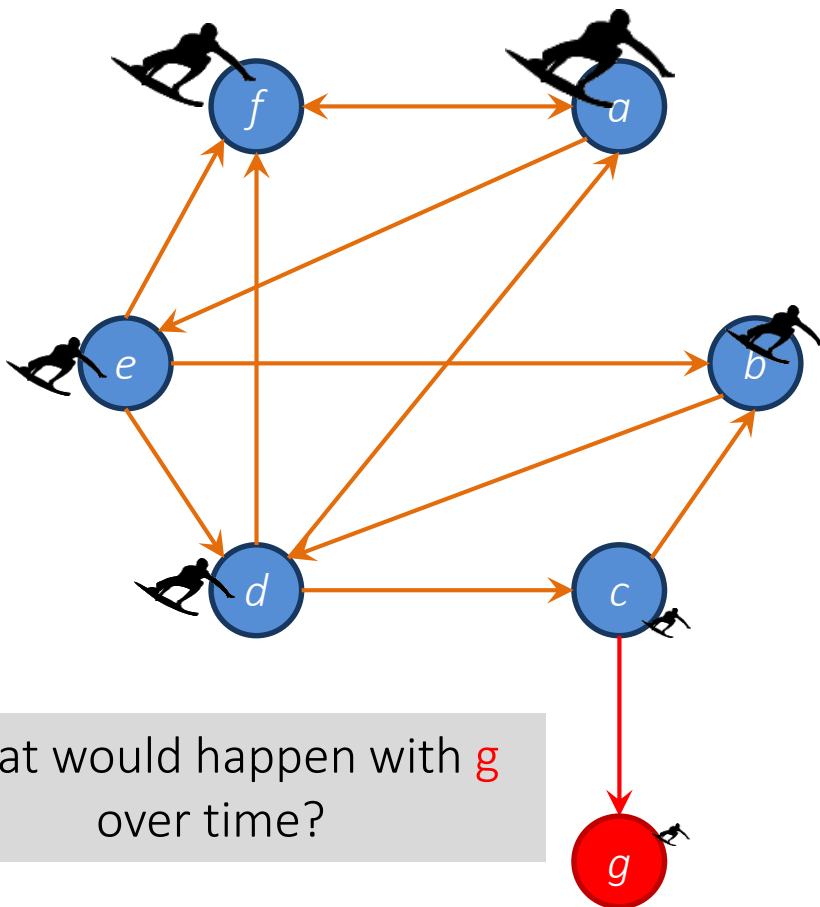


- What is the probability of being at page x after n hops?
- *Initial state*: surfer equally likely to start at any node

PageRank: Random Surfer Model



= someone surfing the web,
clicking links randomly



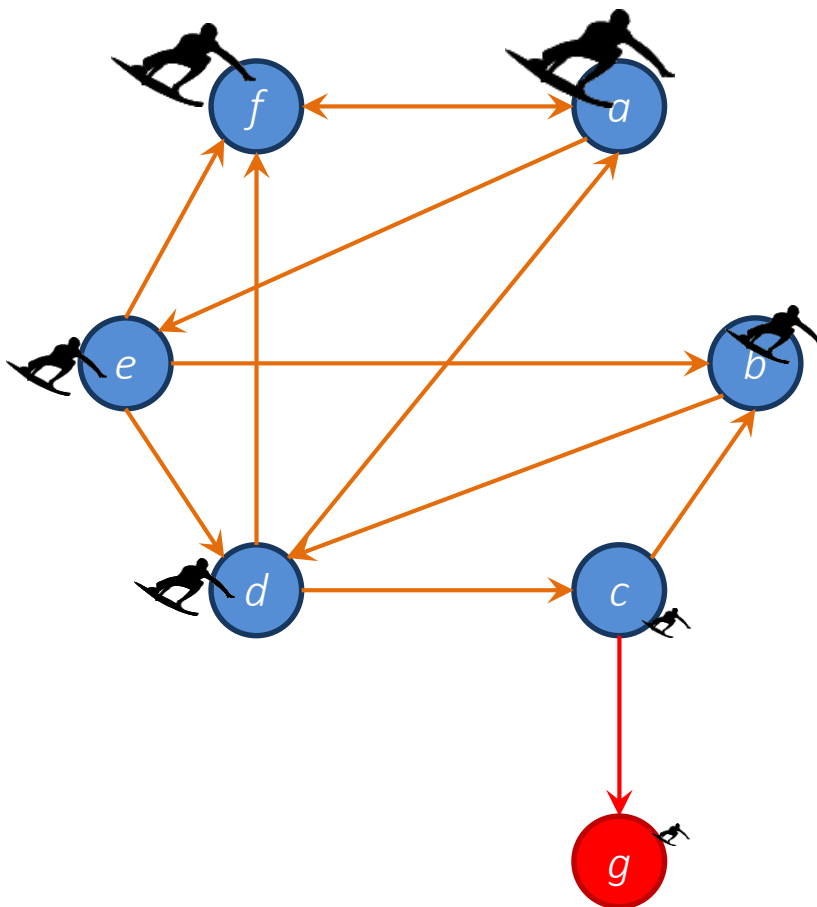
What would happen with **g**
over time?

- What is the probability of being at page x after n hops?
- *Initial state*: surfer equally likely to start at any node
- PageRank applied iteratively for each hop: score indicates probability of being at that page after that many hops

PageRank: Random Surfer Model



= someone surfing the web,
clicking links randomly

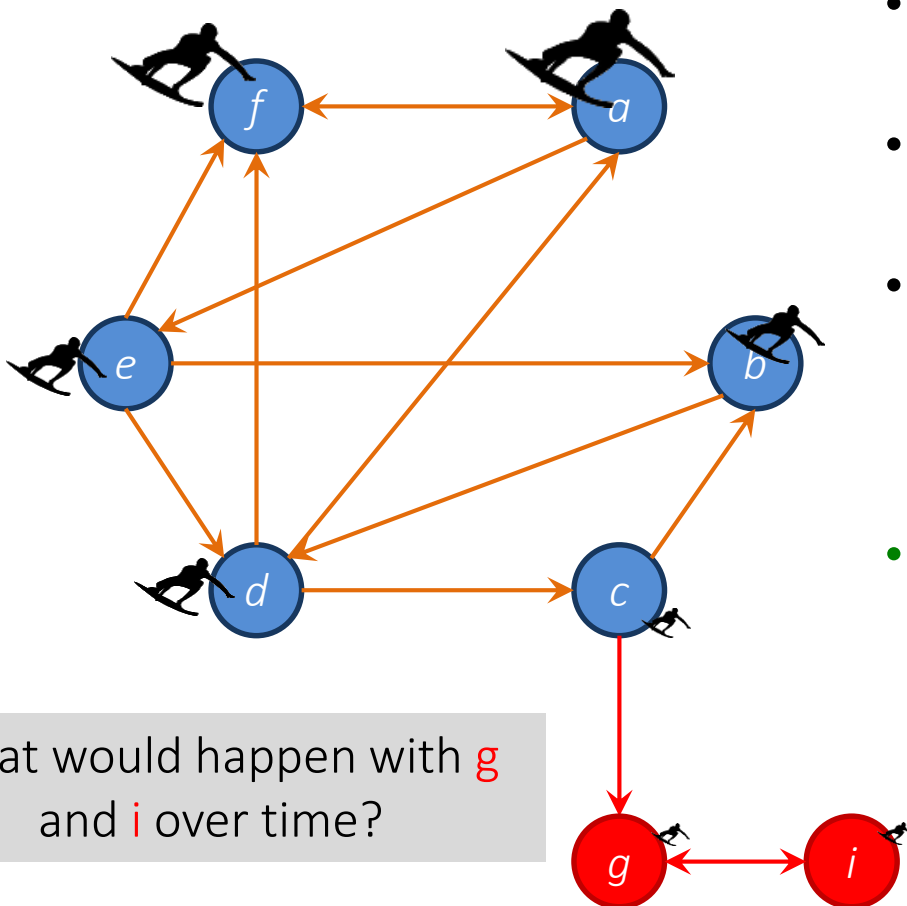


- What is the probability of being at page x after n hops?
- *Initial state*: surfer equally likely to start at any node
- PageRank applied iteratively for each hop: score indicates probability of being at that page after that many hops
- If the surfer reaches a page without links, the surfer randomly jumps to another page

PageRank: Random Surfer Model



= someone surfing the web,
clicking links randomly

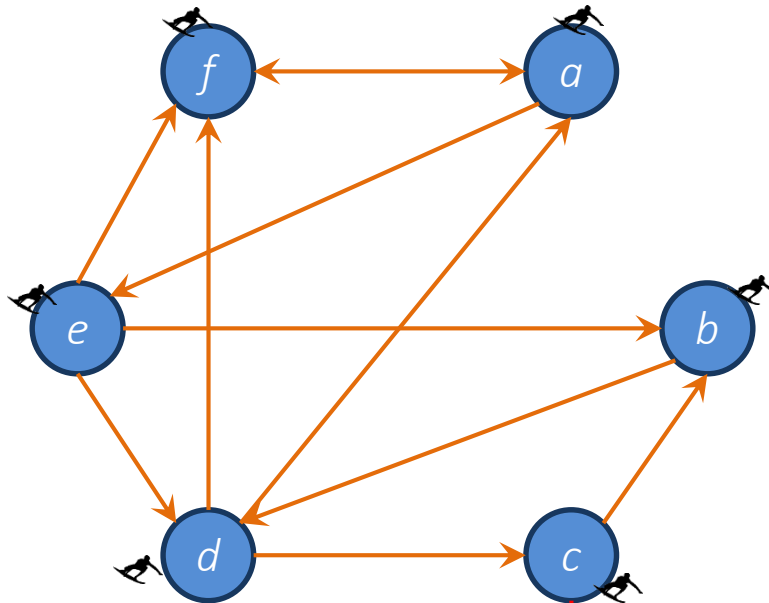


- What is the probability of being at page x after n hops?
- *Initial state*: surfer equally likely to start at any node
- PageRank applied iteratively for each hop: score indicates probability of being at that page after that many hops
- If the surfer reaches a page without links, the surfer randomly jumps to another page

PageRank: Random Surfer Model

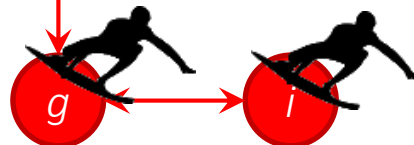


= someone surfing the web,
clicking links randomly



What would happen with **g**
and **i** over time?

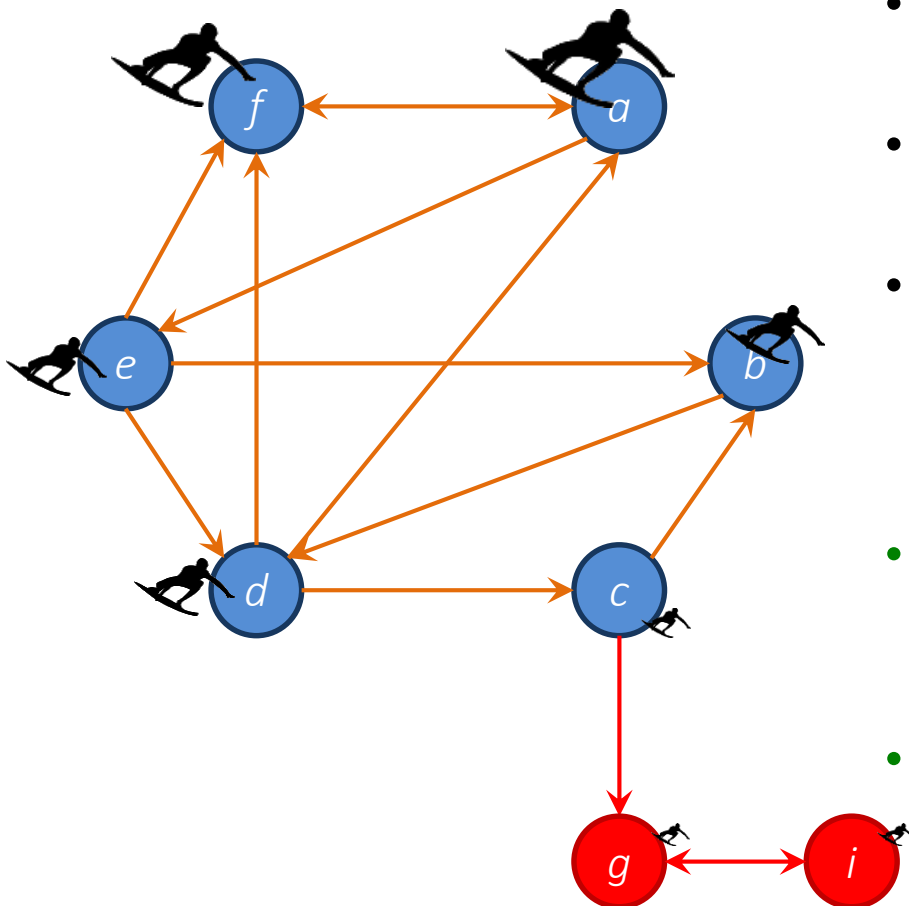
- What is the probability of being at page x after n hops?
- *Initial state*: surfer equally likely to start at any node
- PageRank applied iteratively for each hop: score indicates probability of being at that page after than many hops
- If the surfer reaches a page without out-links, the surfer randomly jumps to another page



PageRank: Random Surfer Model



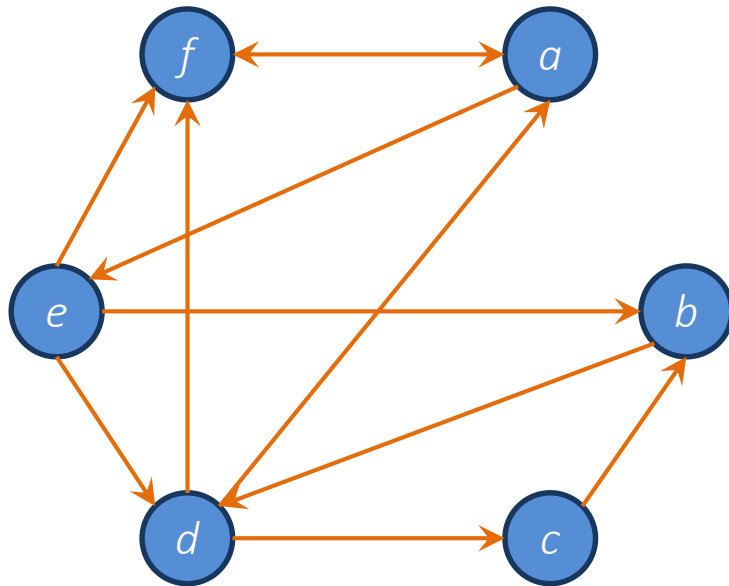
= someone surfing the web,
clicking links randomly



- What is the probability of being at page x after n hops?
- *Initial state*: surfer equally likely to start at any node
- PageRank applied iteratively for each hop: score indicates probability of being at that page after than many hops
- If the surfer reaches a page without out-links, the surfer randomly jumps to another page
- The surfer will jump to a random page at any time with a probability $1 - d$... *this avoids traps and ensures convergence!*

PageRank Model: Final Version

- The Web: a directed graph



$$G = [V, E]$$

Vertices
(pages)

Edges
(links)

$$\text{out}(v) \coloneqq \{v' \in V \mid (v, v') \in E\}$$

$$\text{in}(v) \coloneqq \{v' \in V \mid (v', v) \in E\}$$

$$\text{rank}_0(v) \coloneqq \frac{1}{|V|}$$

$$V' \coloneqq \{v \in V : |\text{out}(v)| = 0\}$$

$$V'' \coloneqq \{v \in V : |\text{out}(v)| \neq 0\}$$

d is the follow-a-link probability
typically ($d = 0.85$)

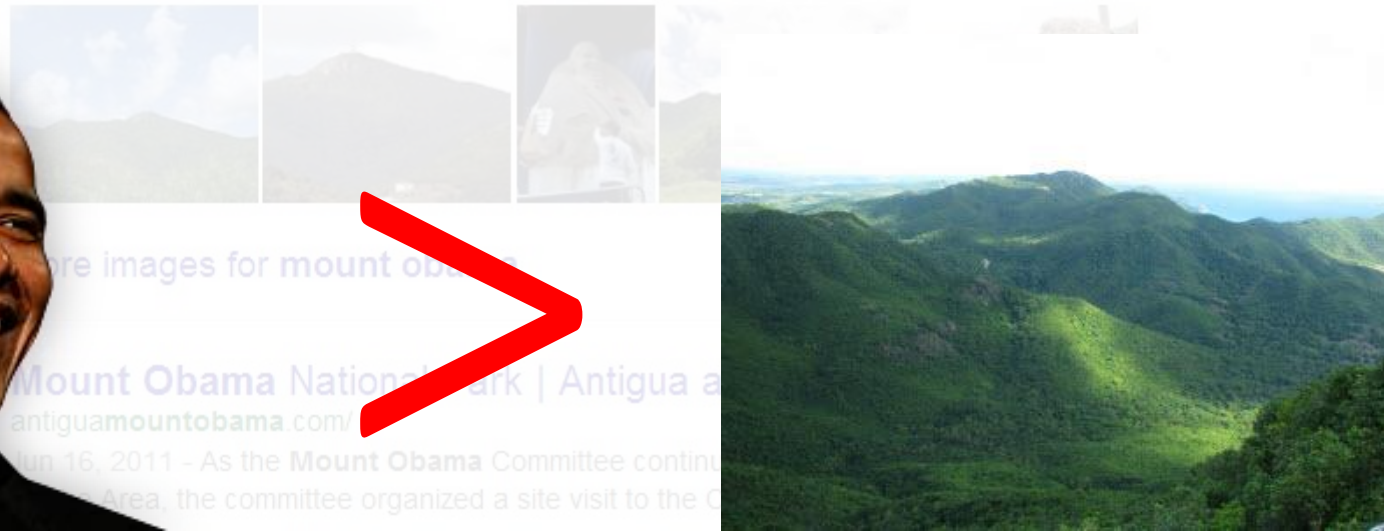
$$\text{rank}_i(v) \coloneqq d \times \sum_{u \in \text{in}(v)} \frac{\text{rank}_{i-1}(u)}{|\text{out}(u)|} + \sum_{v' \in V'} \frac{\text{rank}_{i-1}(v')}{|V|} + (1-d) \times \sum_{v'' \in V''} \frac{\text{rank}_{i-1}(v'')}{|V|}$$

PageRank: Benefits



- ✓ More robust than a simple link count
- ✓ Fewer ties than link counting
- ✓ Scalable to approximate (for sparse graphs)
- ✓ Convergence guaranteed

Two Sides to Ranking: Importance

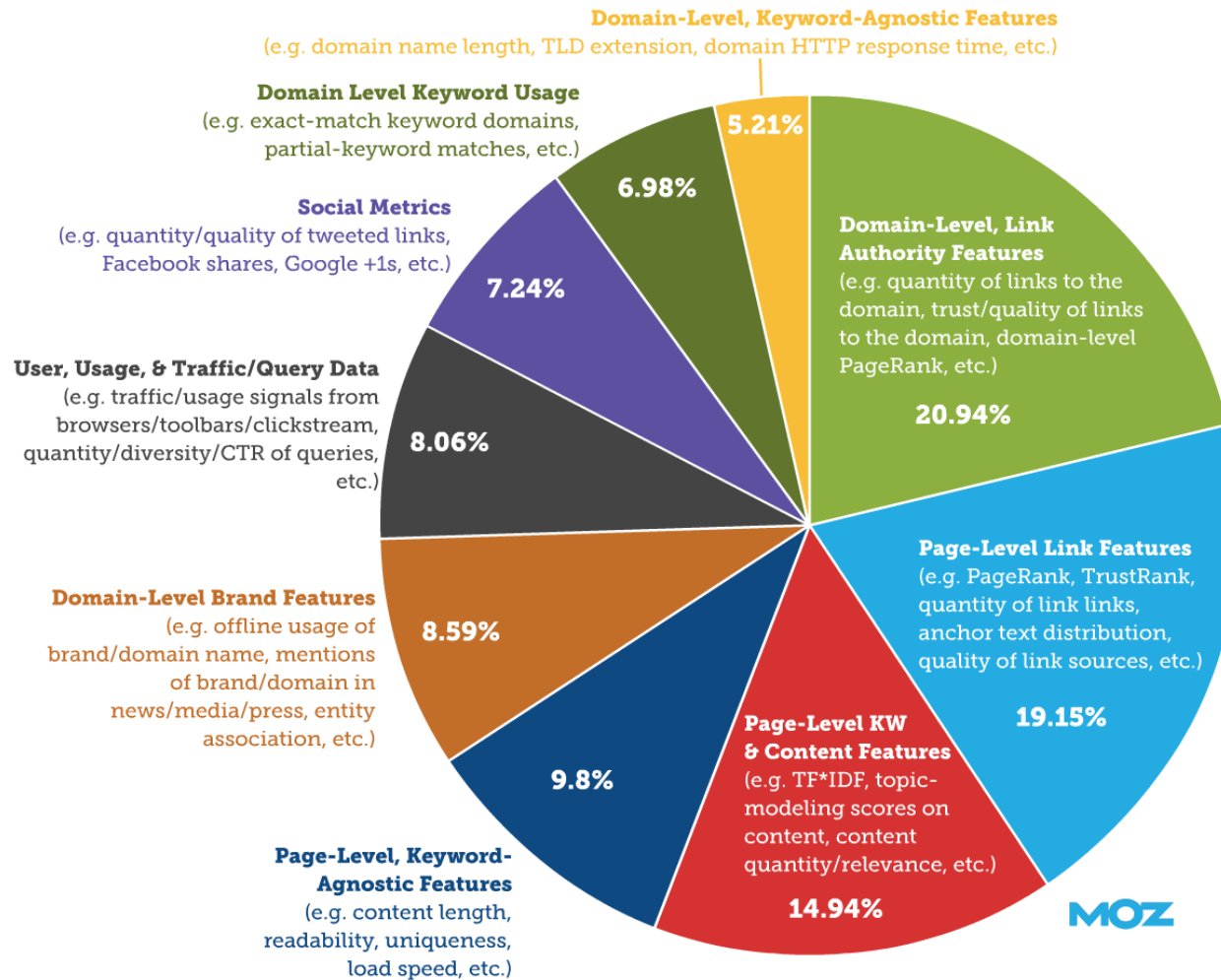


HOW DOES GOOGLE REALLY RANK?
AN EDUCATED GUESS

How Modern Google ranks results (maybe)

Weighting of Thematic Clusters of Ranking Factors in Google

(based on survey responses by 128 SEO professionals in June 2013)

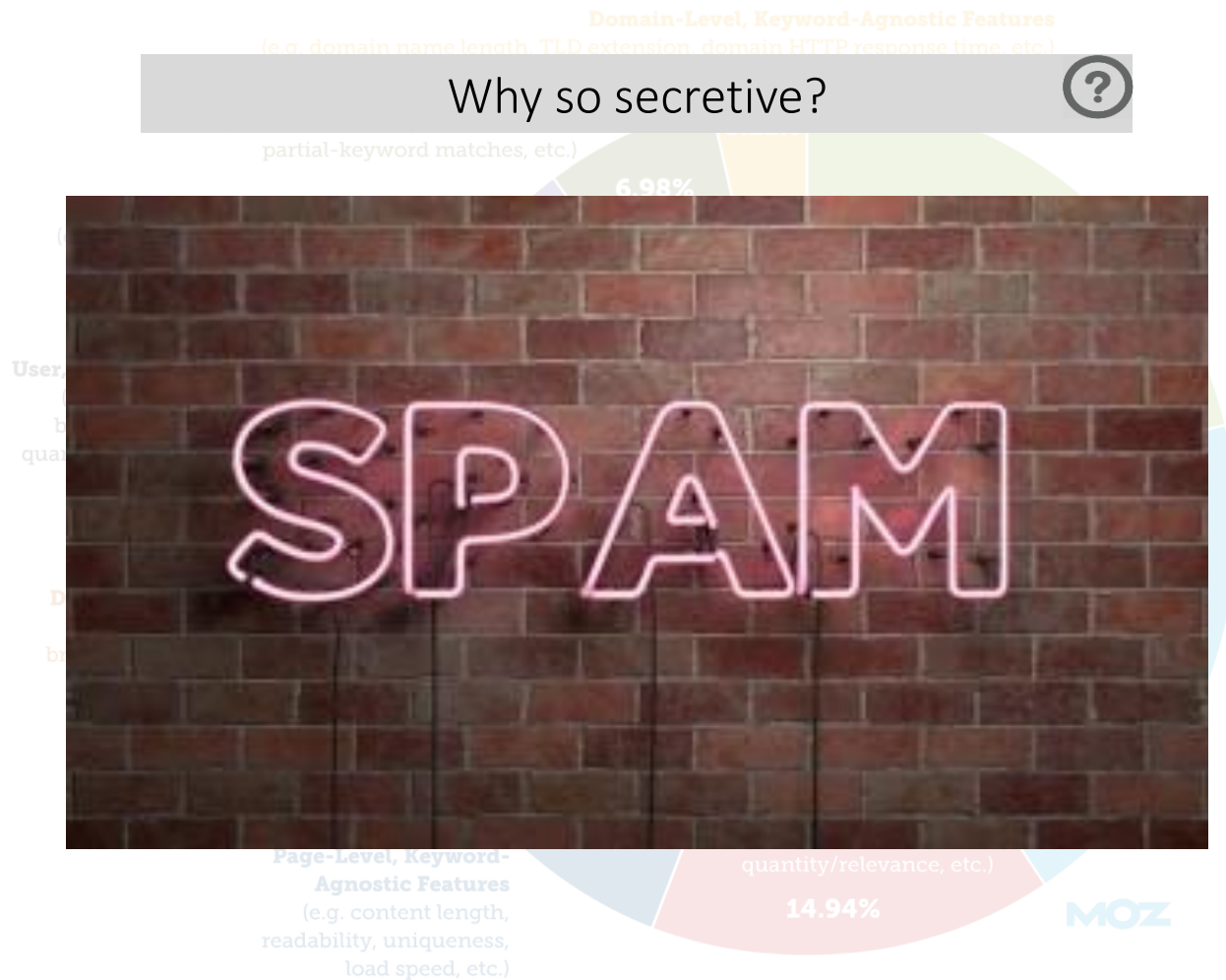


According to survey of SEO experts, not people in Google

How Modern Google ranks results (maybe)

Weighting of Thematic Clusters of Ranking Factors in Google

(based on survey responses by 128 SEO professionals in June 2013)



According to survey of SEO experts, not people in Google

Ranking: Science or Art?



