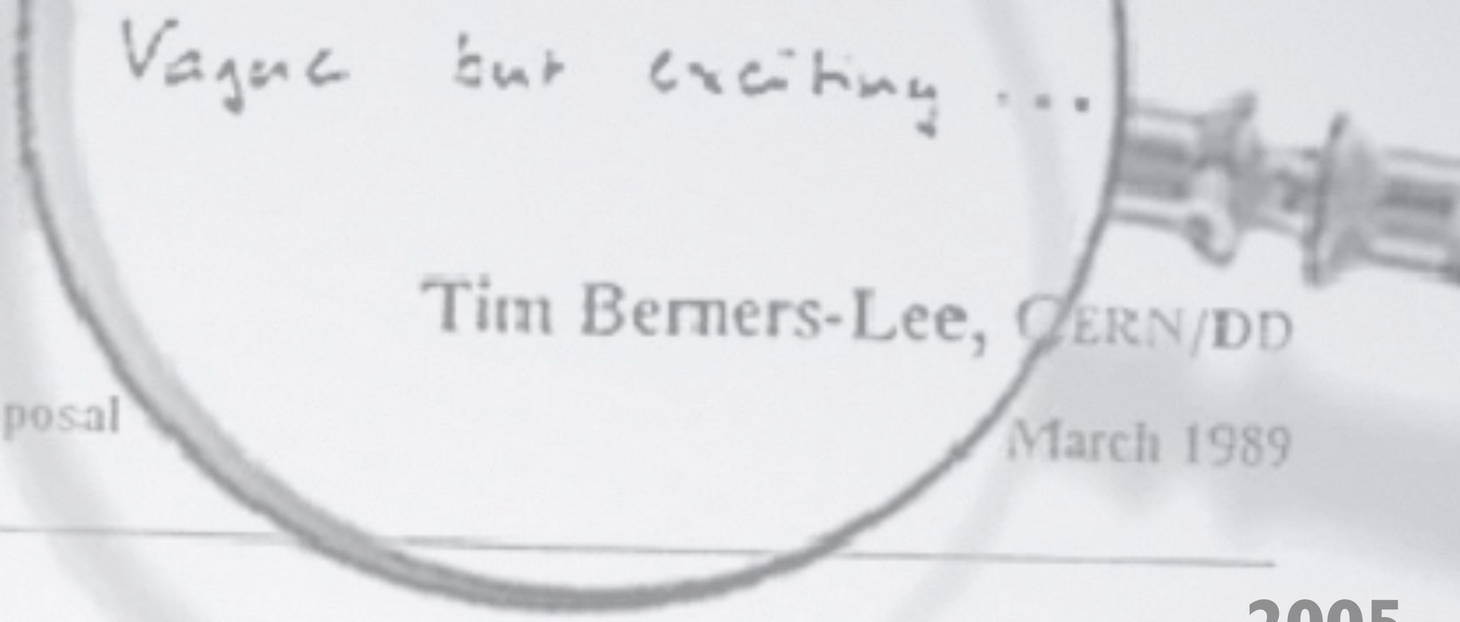
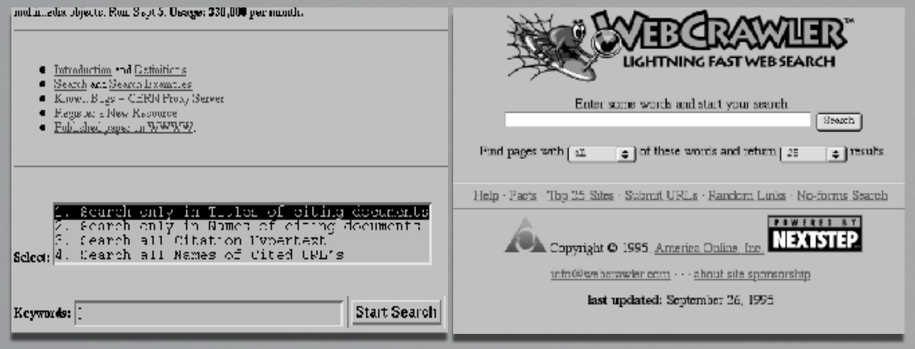
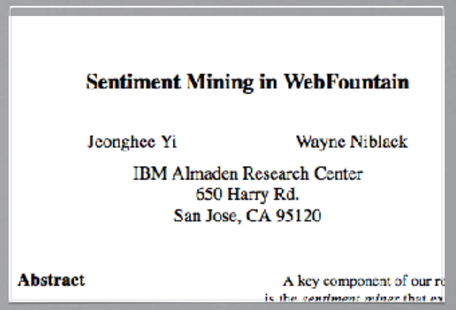

W3C
conferência
web.br
2017



1993 — 1994 — 1997 — 1998 — 2001 — 2005





The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page
{sergey, page}@cs.stanford.edu
Computer Science Department, Stanford University, Stanford, CA 94305

Abstract

We assume page A has pages T1...Tn which point to it (i.e., are citations). The parameter d is a damping factor which can be set between 0 and 1. We usually set d to 0.85. There are more details about d in the next section. Also C(A) is defined as the number of links going out of page A. The PageRank of a page A is given as follows:

$$PR(A) = (1-d) + d\,(PR(T1)/C(T1) + \ldots + PR(Tn)/C(Tn))$$

Note that the PageRanks form a probability distribution over web pages, so the sum of all web pages' PageRanks will be one.

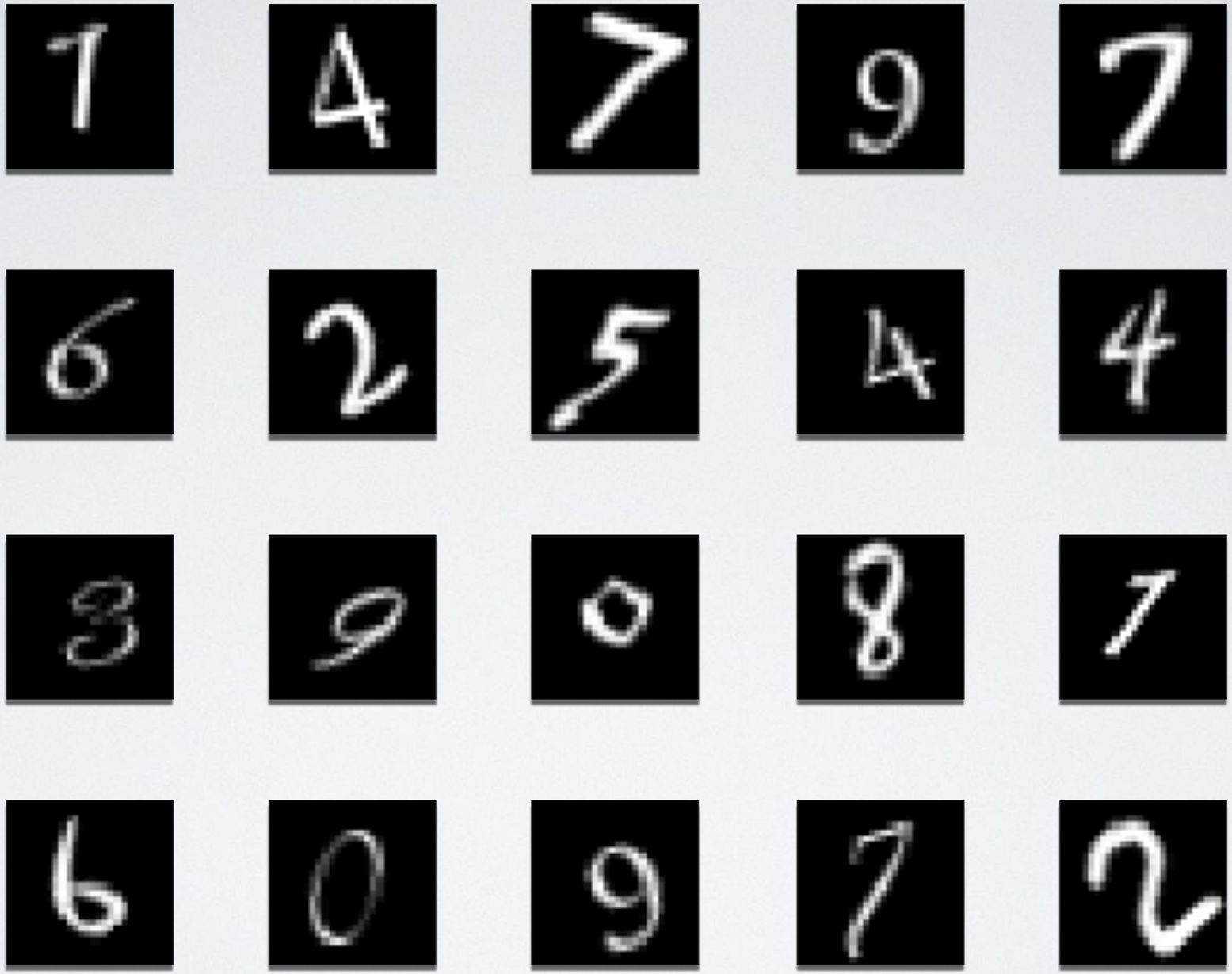
Verification of a human in the loop or Identification via the Turing Test*

Moni Naor †

September 13th, 1996

Abstract

We propose using a "Turing Test" in order to verify that a human is the one making a query to a service over the web. Thus, before a request is processed the user should answer as a challenge an instance of a problem chosen so that it is easy for humans to solve but the best known programs fail on a non-negligible fraction of the instances. We discuss several scenarios where such tests are desired and several potential sources for problems instances. We also dicuss the application of this idea for combatting junk mail.



```
URL="http://www2.planalto.gov.br/acompanhe-planalto/"\
    "discursos/discursos-do-presidente-da-republica/"\
    "discurso-do-presidente-da-republica-michel-temer-"\
    "durante-cerimonia-de-encerramento-do-ii-encontro-"\
    "da-carta-caiman"
```

```
import requests
from selenium import webdriver
```

```
r=requests.get(URL, timeout=20)
```

```
timeout                                                  Traceba
```

```
In [6]: wd=webdriver.PhantomJS()

In [7]: wd.get(URL)

In [8]: wd.save_screenshot("govTest.png")
Out[8]: True
```

Apache Nutch

Heritrix

Scrapy

PhantomJS

Selenium


jobs eating more international edition
theguardian
tech travel all sections
How the internet found a better way than illegible squiggles to prove you're not a robot
Captcha has evolved from identifying mangled letters to web users unwittingly training Google's AI. Now, finally, you won't have to do anything
I'm not a robot
reCAPTCHA


NEW! The Free & Easy Way to Prevent Bots from Ruining Google Analytics
Are You a Human is now part of Distil Networks
Read the Announcement
Get Bot-Filtered Analytics
Bots 45,289
62%
38%
Humans 117,940
SHIELDSQUARE
TECHNOLOGY
HOW IT WORKS
INTEGRATION OPTIONS
WEB CRAWLER BLOCKING SERVICE
Detect and Stop Suspicious Crawling Activities on Your Website With ShieldSquare
Protecting billions of Web pages for businesses spread across 70 countries

Information extraction — or automatically classifying data items stored as plain text — is a major topic of artificial-intelligence research.

Image: MIT News

Artificial-intelligence system surfs web to improve its performance

"Information extraction" system helps turn plain text into data for statistical analysis.

Larry Hardesty | MIT News Office
November 10, 2016

Press Inquiries

RELATED

Paper: "Improving information extraction by

Of the vast wealth of information unlocked by the Internet, most is plain text. The data

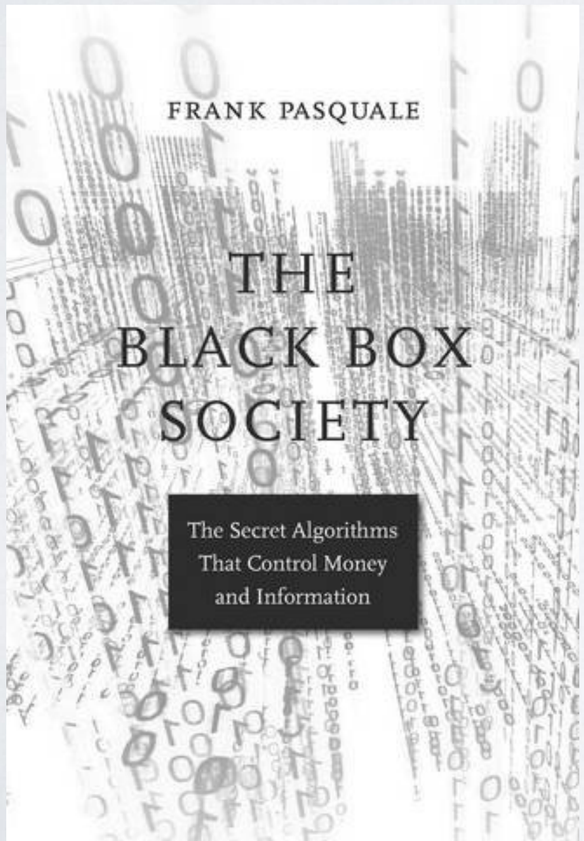
Text Generation With LSTM Recurrent Networks in Python with Keras

by **Jason Brownlee** on August 4, 2016 in **Long Short-Term Memory Networks**

Recurrent neural networks can also be used as generative models.

This means that in addition to being used for predictive models (making predictions the sequences of a problem and then generate entirely new plausible sequences f domain.

Generative models like this are useful not only to study how well a model has learn to learn more about the problem domain itself.



TOM SIMONITE BUSINESS 10.18.17 03:00 PM

AI EXPERTS WANT TO END 'BLACK BOX' ALGORITHMS IN GOVERNMENT

Intelligent Machines

The Dark Secret at the Heart of AI

No one really knows how the most advanced algorithms do what they do. That could be a problem.

by Will Knight April 11, 2017

Last year, a strange self-driving car was released onto the quiet roads of Monmouth County, New Jersey. The experimental vehicle, developed by researchers at the chip maker Nvidia, didn't look different from other autonomous cars, but it was unlike anything demonstrated by Google, Tesla, or General Motors, and it showed the rising power of artificial intelligence. The car didn't follow a single instruction provided by an engineer or programmer. Instead, it

Obrigado

Rodrigo Arrigoni

Twitter e Github: @VulcanoAhab