

# Internet-in-a-Box Project

*Sharing the world's Free information*

*Presented by*

**Braddock Gaskill**

Humane Informatics LLC

SGVHAK

*11 April 2013*

*SGVLUG*

# What is an Internet-in-a-Box?





# What is an Internet-in-a-Box?

- a small, inexpensive device (\$250)
- essential Internet resources without any Internet connection.
- a local copy of almost a terabyte of the world's Free information.

# Resources

**Wikipedia:** 40 languages

**Maps:** world-wide to street level

**E-books:** 40,000 books in many languages

**Software:** Open Source. Linux OS with all software package repositories. Source code.

**Video:** 500+ hours of instruction

# For Who?

For schools and communities in the developing world.



# Why?

- Internet access is still expensive and difficult in the developing world
- Inexpensive media now exists that can hold vast offline libraries of information
- Large sources of content are available
- Inexpensive reader devices now exist which can read libraries on media

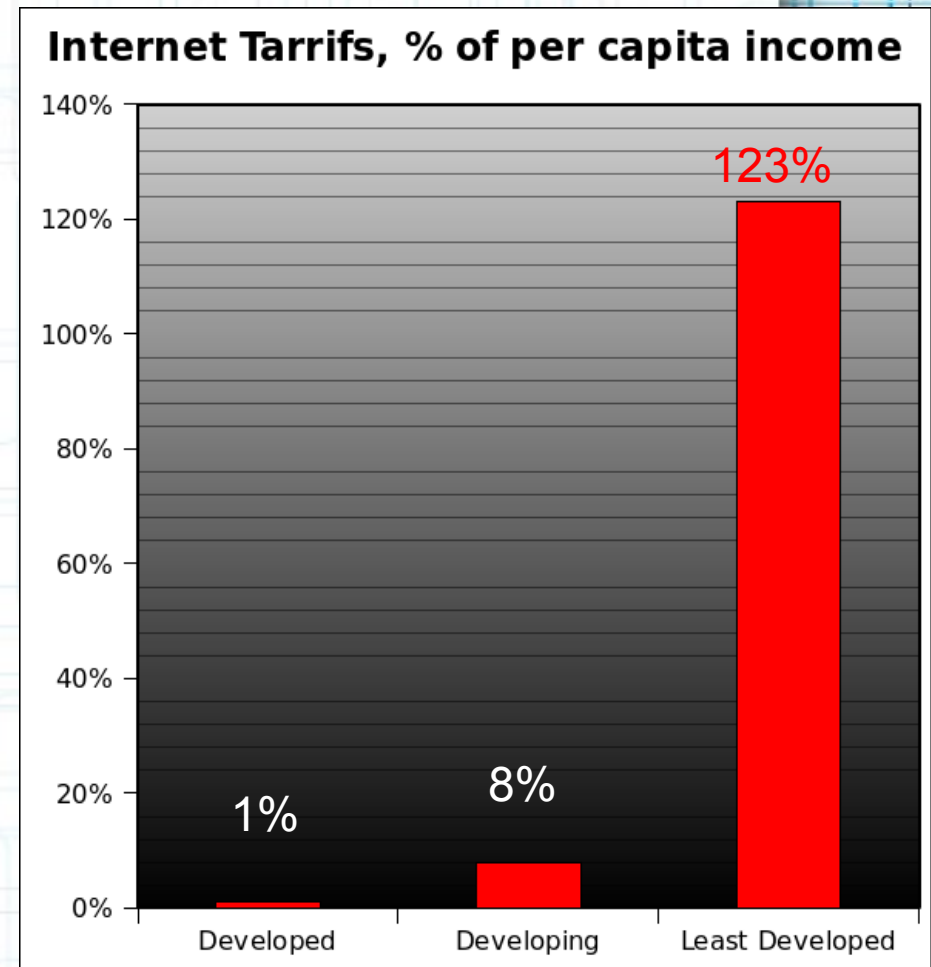
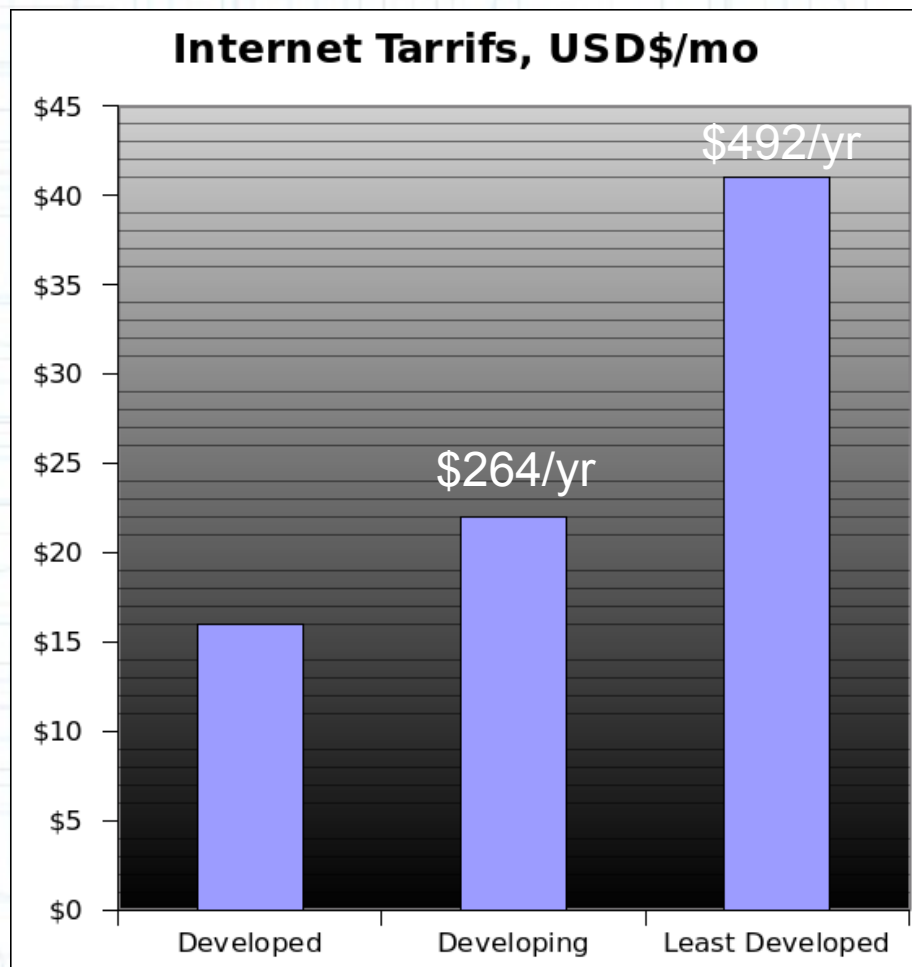
# Internet Penetration

- **70%** use internet in "**developed**" countries
- Only **24%** use internet in "**developing**" countries\*

\* 2013, Napoli & Obar

# Internet Cost

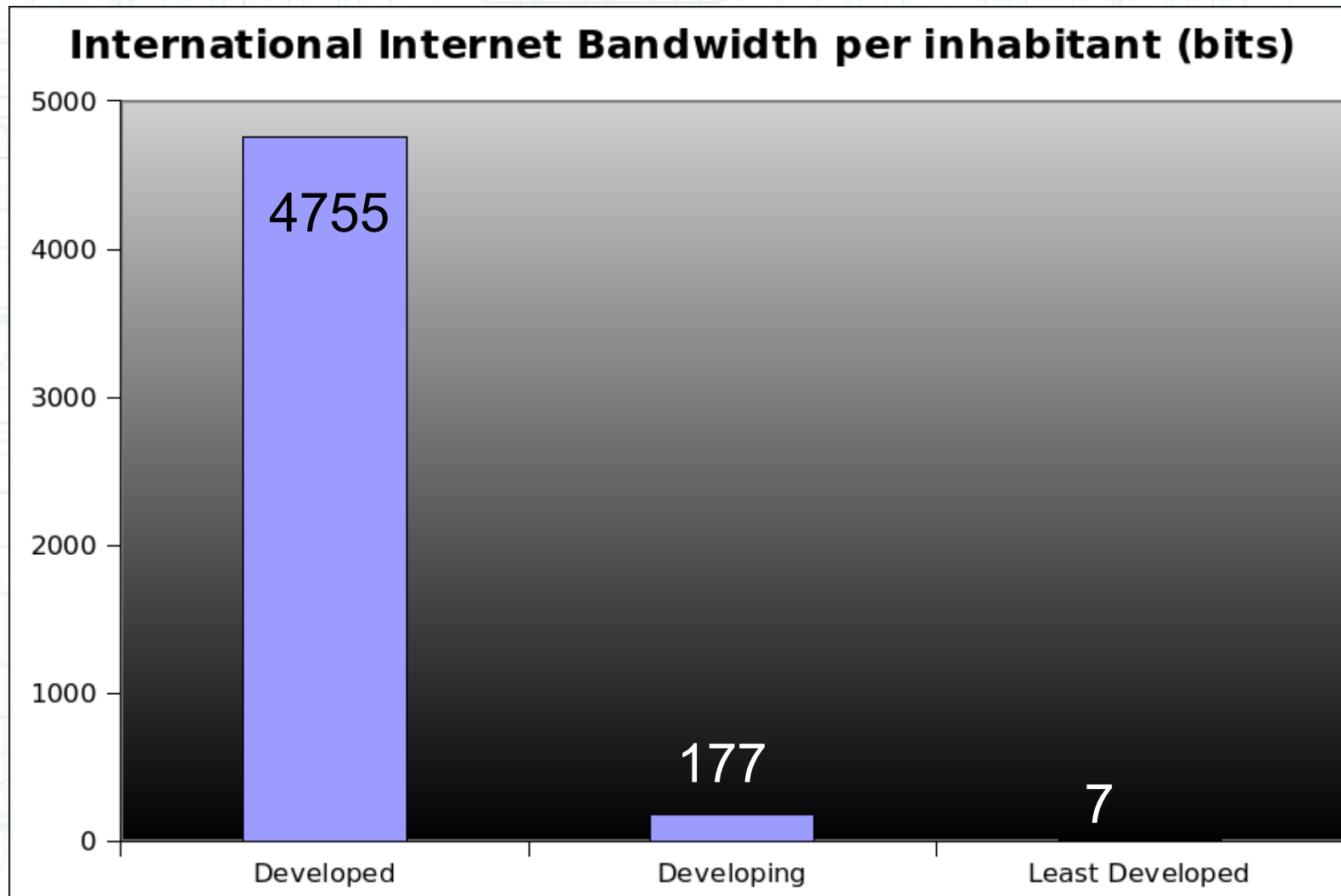
- Cost for 20 hours of dialup per month





# Internet Capacity

- Deficit of internet infrastructure



# We Have The Technology

- **Half of the US Library of Congress Printed Collection would fit on ONE hard drive (text only)**



# The Law

You are **NOT** allowed to share most information



# What can we share?

- Open Source
  - Explicitly licensed for sharing
- Public Domain
  - Generally, pre-1923 is Free
  - Congress often retro-actively extending.  
New material from 1923 starting in 2019  
IF nothing changes.
  - Some exceptions, it's complicated

# Wired and Wireless

## Wireless

- Wifi Hotspot
- Battery
- Solar option



## Wired

- Connect to existing network
- Network appliance



# User Devices

**One Laptop  
Per Child**  
2.5 Million  
users



**Tablets**  
\$50 in India

**Phones**  
with wifi

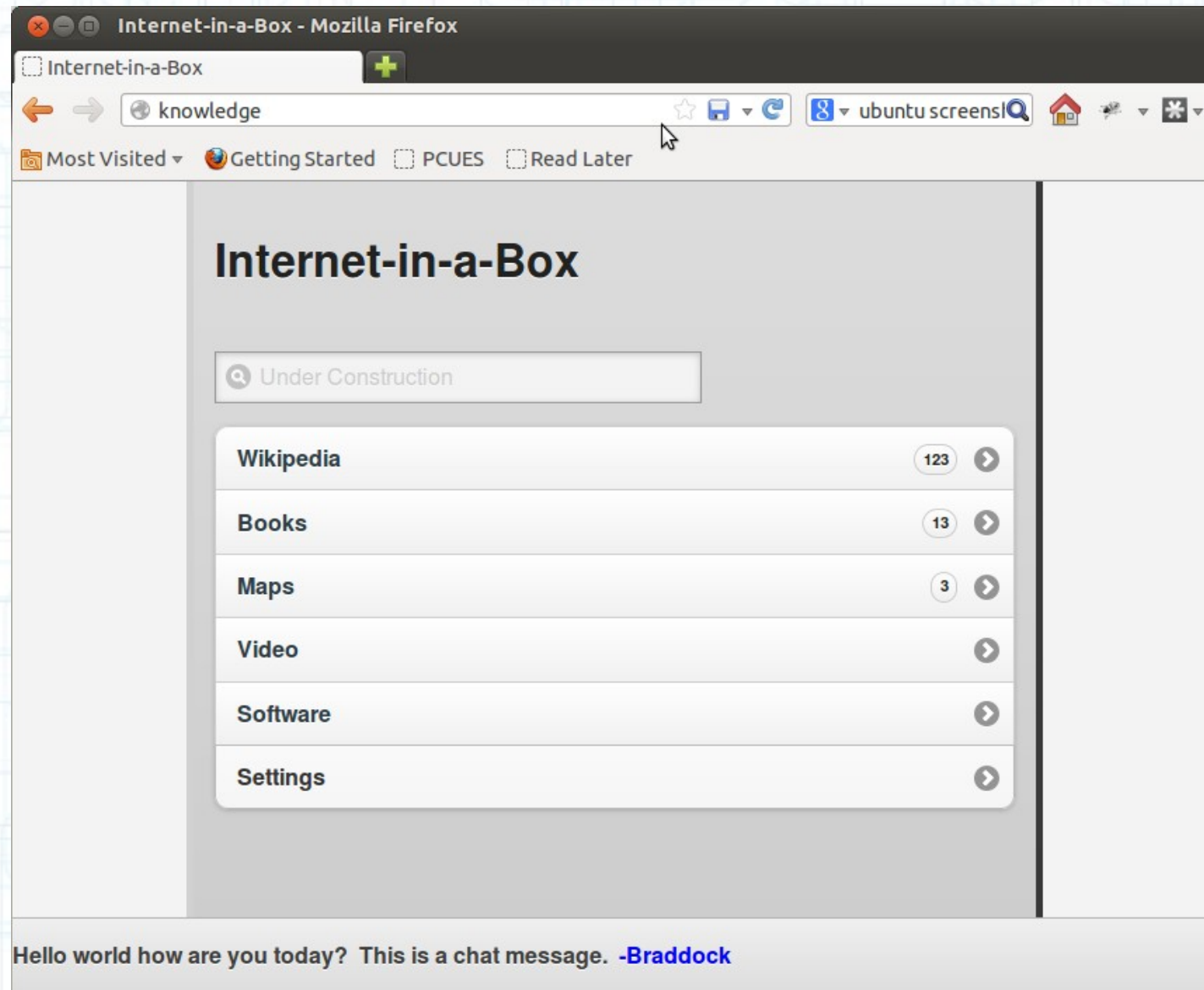




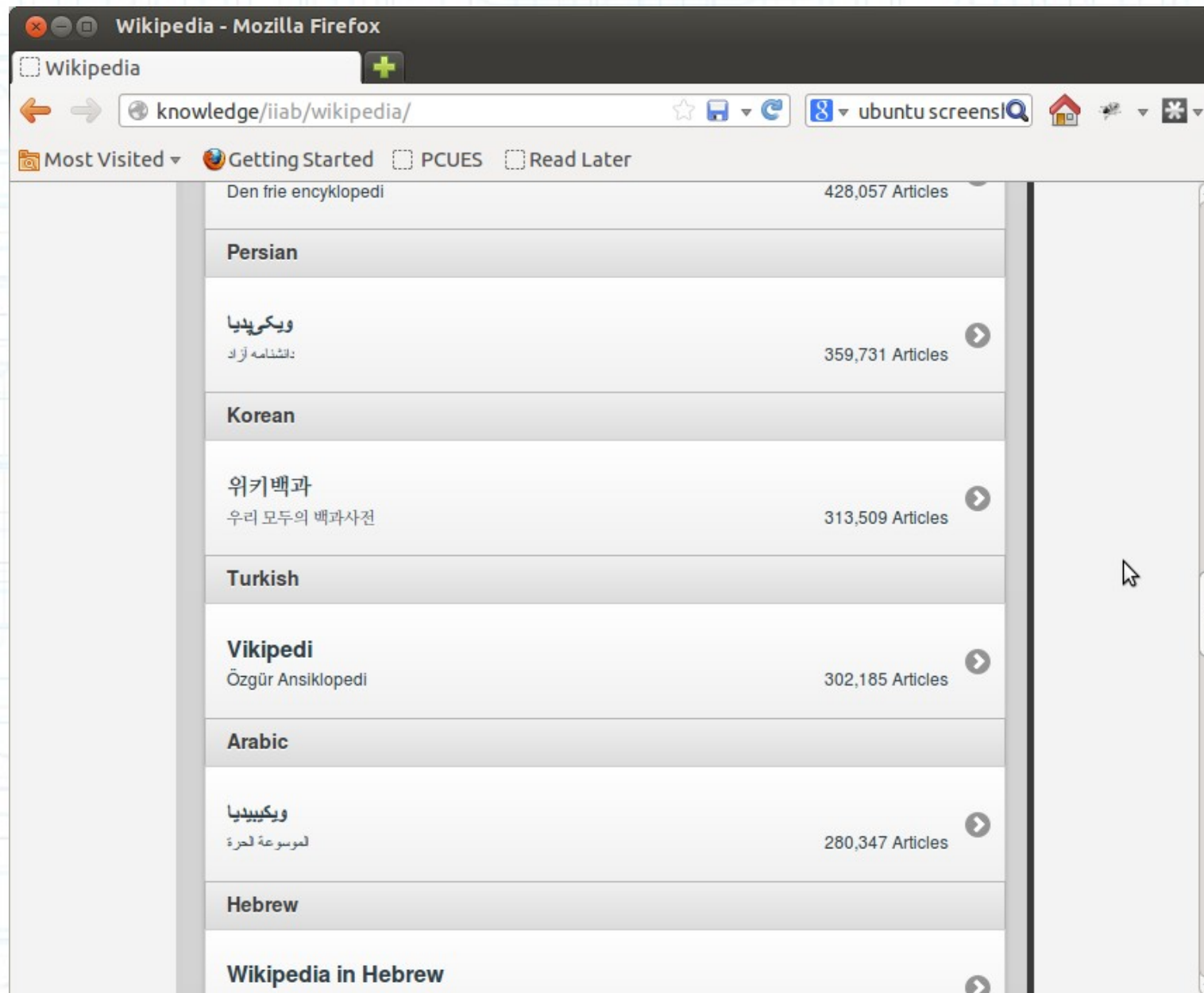


**Demo**

# Main Menu



# Wikipedia





# Wikipedia Article

위키백과 - Mozilla Firefox

위키백과

192.168.1.113:25001/wikipedia\_ko\_all\_10\_2012/

ubuntu screens

Most Visited Getting Started PCUES Read Later

Library Home Search

## 위키백과

**위키백과**(듣기 (도움말·정보), Wiki百科, 영어: Wikipedia 위키피디아<sup>[\*]</sup> 듣기 (도움말·정보)) 혹은 **위키피디아**는 모두가 함께 만들어 가며 누구나 자유롭게 쓸 수 있는 다언어판 **인터넷 백과사전**이다. 대표적인 **집단 지성**의 사례로 평가 받고 있다. 배타적인 **저작권**을 가지고 있지 않기 때문에 사용에 제약을 받지 않는다.

2001년 1월 15일에 시작된 위키백과는 비영리 단체인 위키미디어 재단에서 운영하고 있다. 2012년 3월 현재 **영어판** 388만여 개, **한국어판** 20만여 개를 비롯하여 280여 언어판을 합하면 2100만여개의 글이 수록되어 있으며 꾸준히 성장하고 있다.


위키백과 **영어판**은 전문가들이 작성했던 백과사전인 **누피디어**(지금은 없어짐)를 보완하여 2001년 1월 15일에 만들어졌다. **지미 웨일스**와 **래리 생거**는 2001년에 위키백과의 공동창업자로 인정받았다.

목차 [숨기기]


1 역사

2 운영

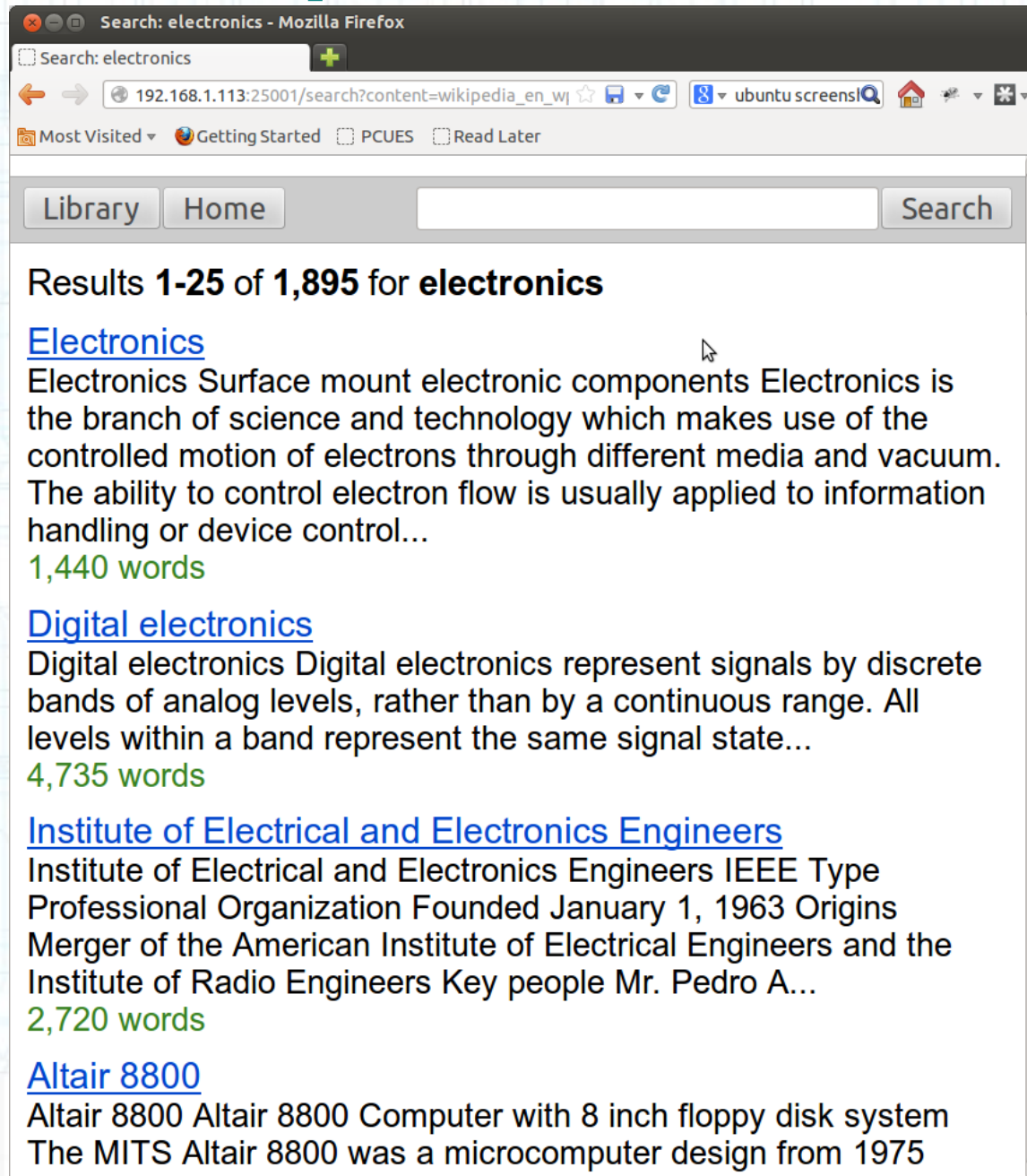
2.1 위키미디어 재단과 위키아



위키백과의 공식 로고

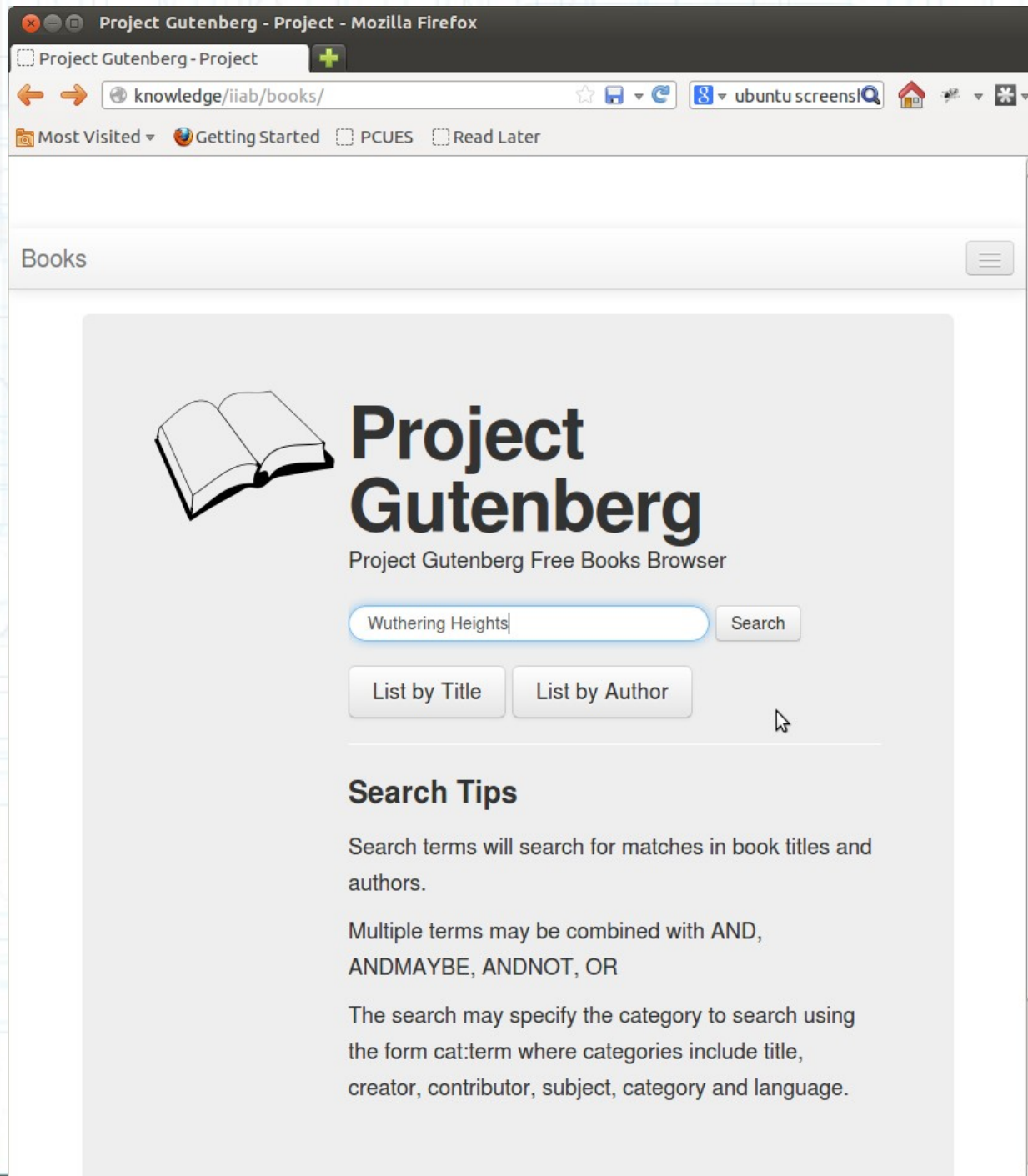


# Wikipedia Search





# Books





# Books

Book Details - Project - Mozilla Firefox

Book Details - Project

knowledge/iiab/books/text/etext768/details

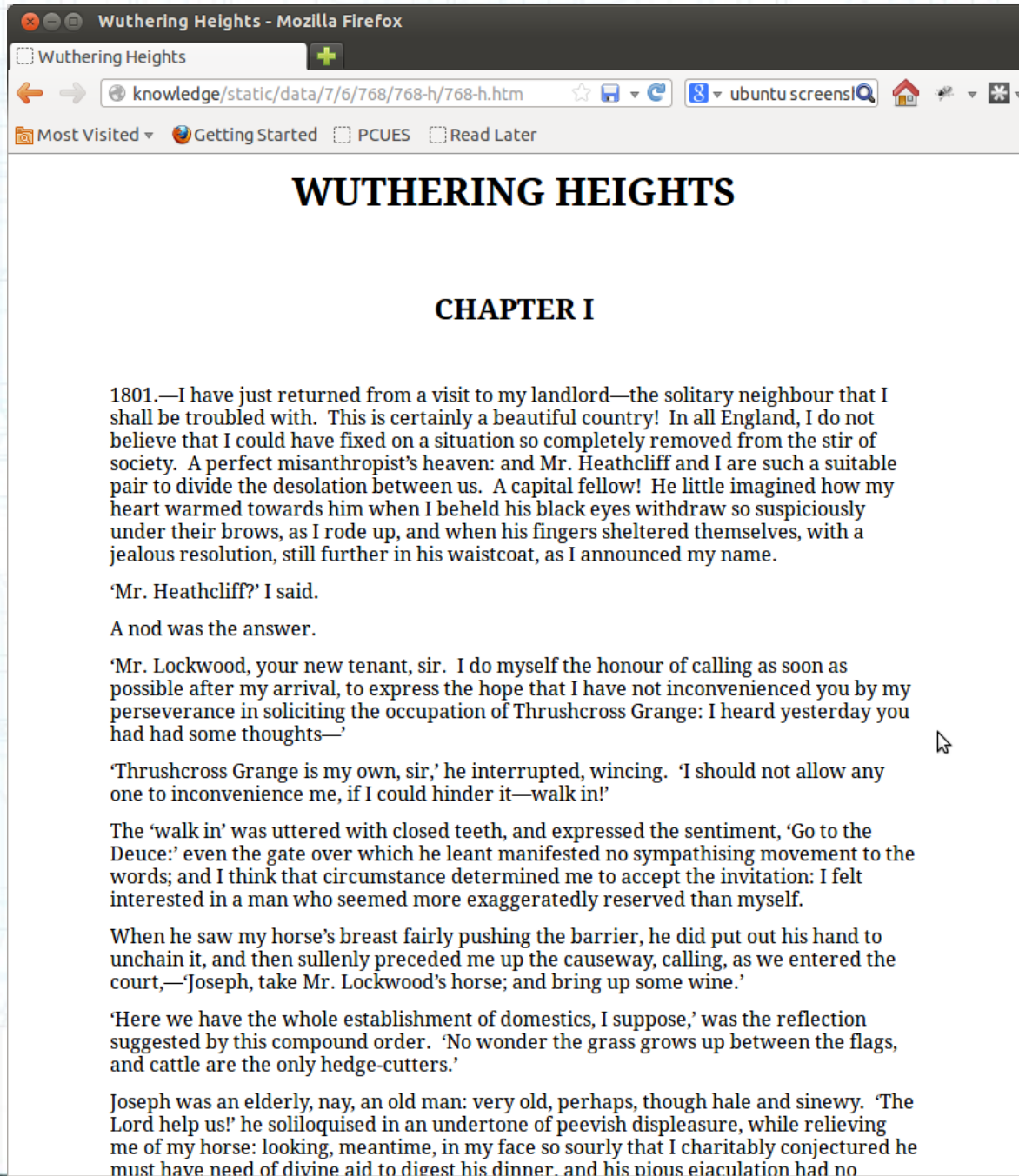
Most Visited Getting Started PCUES Read Later

Books

<b>Title</b>	Wuthering Heights
<b>Author</b>	Brontë, Emily, 1818-1848
<b>Subject</b>	PR
<b>Subject</b>	England -- Social life and customs -- 19th century -- Fiction
<b>Subject</b>	Love stories
<b>Subject</b>	Domestic fiction
<b>Subject</b>	Psychological fiction
<b>Subject</b>	Foundlings -- Fiction
<b>Subject</b>	Triangles (Interpersonal relations) -- Fiction
<b>Subject</b>	Rejection (Psychology) -- Fiction
<b>Subject</b>	Yorkshire (England) -- Fiction
<b>Subject</b>	Rural families -- Fiction
<b>Subject</b>	Heathcliff (Fictitious character) -- Fiction
<b>Language</b>	en

Name	Format
<a href="#">data/7/6/768/768-h/768-h.htm</a>	text/html
<a href="#">data/7/6/768/768.txt</a>	text/plain; charset="us-ascii"
<a href="#">data/cache/generated/768/pg768.epub</a>	application/epub+zip

# Read



The screenshot shows a Mozilla Firefox browser window with the title 'Wuthering Heights - Mozilla Firefox'. The address bar shows the URL 'knowledge/static/data/7/6/768/768-h/768-h.htm'. The page content is titled 'WUTHERING HEIGHTS' and 'CHAPTER I'. The text begins with '1801.—I have just returned from a visit to my landlord—the solitary neighbour that I shall be troubled with. This is certainly a beautiful country! In all England, I do not believe that I could have fixed on a situation so completely removed from the stir of society. A perfect misanthropist's heaven: and Mr. Heathcliff and I are such a suitable pair to divide the desolation between us. A capital fellow! He little imagined how my heart warmed towards him when I beheld his black eyes withdraw so suspiciously under their brows, as I rode up, and when his fingers sheltered themselves, with a jealous resolution, still further in his waistcoat, as I announced my name.'

'Mr. Heathcliff?' I said.

A nod was the answer.

'Mr. Lockwood, your new tenant, sir. I do myself the honour of calling as soon as possible after my arrival, to express the hope that I have not inconvenienced you by my perseverance in soliciting the occupation of Thrushcross Grange: I heard yesterday you had had some thoughts—'

'Thrushcross Grange is my own, sir,' he interrupted, wincing. 'I should not allow any one to inconvenience me, if I could hinder it—walk in!'

The 'walk in' was uttered with closed teeth, and expressed the sentiment, 'Go to the Deuce:' even the gate over which he leant manifested no sympathising movement to the words; and I think that circumstance determined me to accept the invitation: I felt interested in a man who seemed more exaggeratedly reserved than myself.

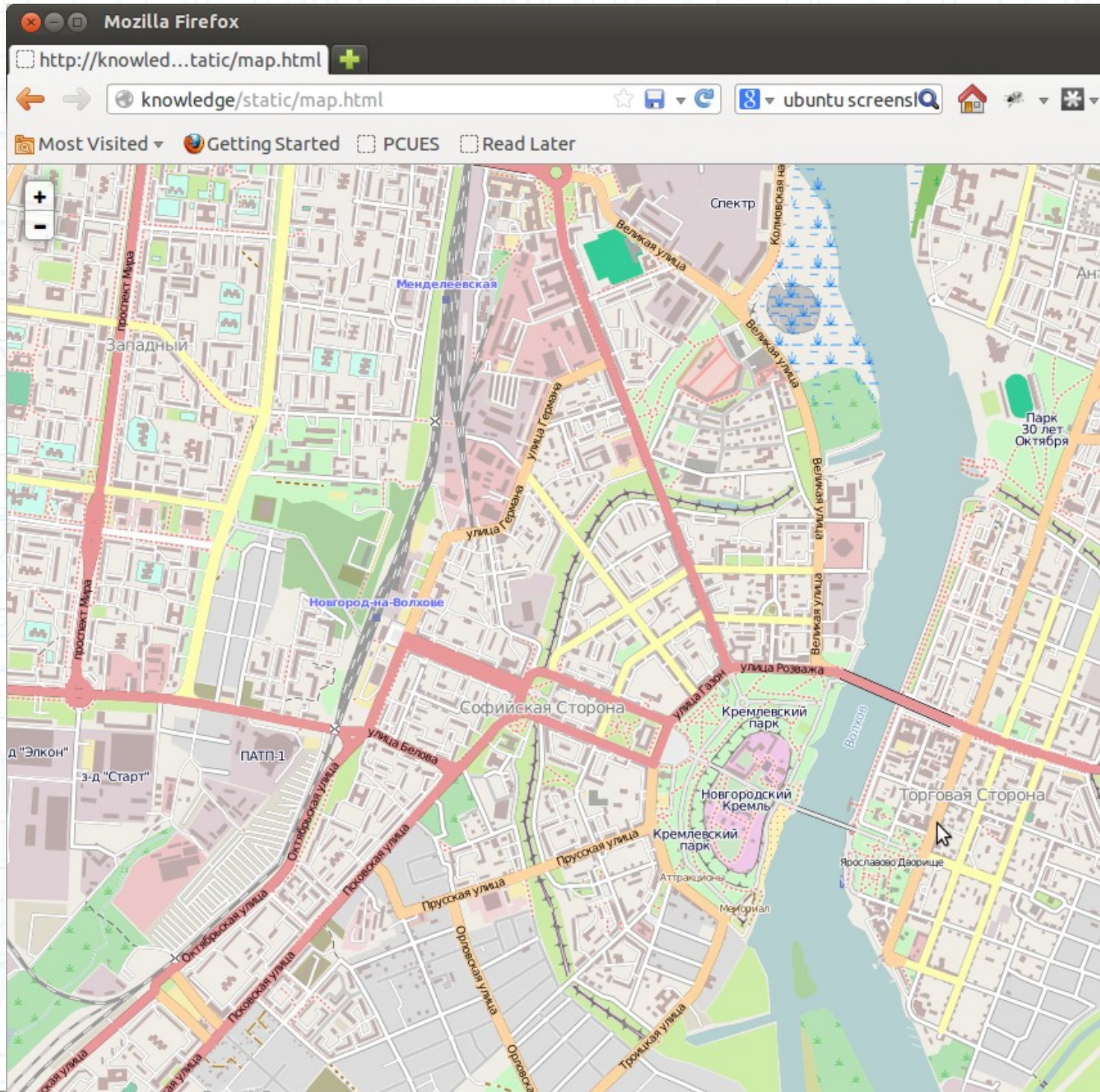
When he saw my horse's breast fairly pushing the barrier, he did put out his hand to unchain it, and then sullenly preceded me up the causeway, calling, as we entered the court,—'Joseph, take Mr. Lockwood's horse; and bring up some wine.'

'Here we have the whole establishment of domestics, I suppose,' was the reflection suggested by this compound order. 'No wonder the grass grows up between the flags, and cattle are the only hedge-cutters.'

Joseph was an elderly, nay, an old man: very old, perhaps, though hale and sinewy. 'The Lord help us!' he soliloquised in an undertone of peevish displeasure, while relieving me of my horse: looking, meantime, in my face so sourly that I charitably conjectured he must have need of divine aid to digest his dinner. and his pious ejaculations had no

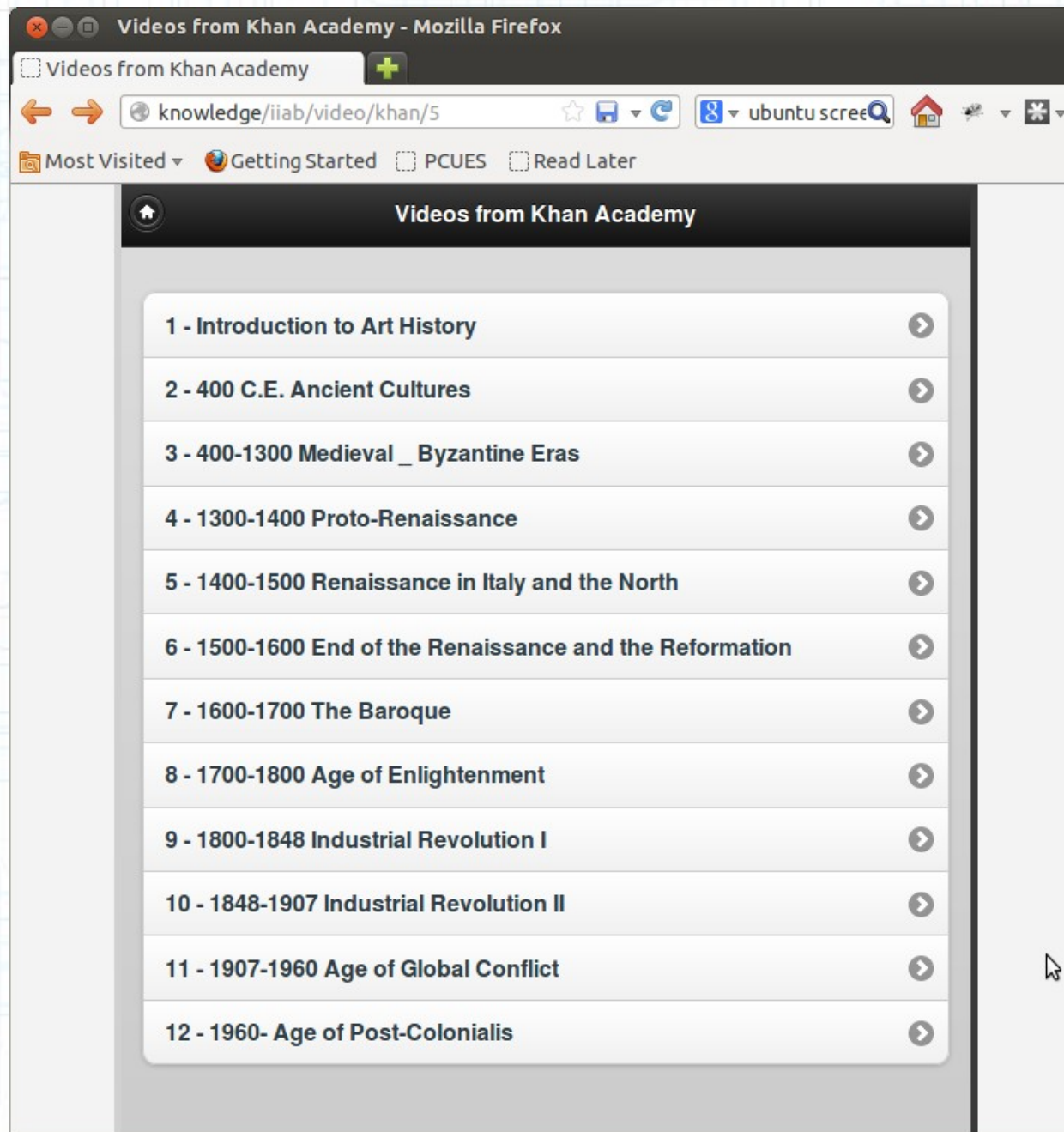


# Maps





# Video



# Video

Videos from Khan Academy - Mozilla Firefox

Videos from Khan Academy

knowledge/iiab/video/khan/5/1/1/1

Most Visited ▾ Getting Started PCUES Read Later

Art History > Introduction to Art History > The Basics > Why Look at Art

1763 Treaty of Paris (1763) X

1765 Stamp Act

1773 Boston Tea Party

1776 Declaration of Independence

Revolutionary War

1783 Treaty of Paris (1783)

1787 - created  
1788 - ratified  
1789 - in effect } Constitution

Washington's two terms as President

1797

1803 Louisiana Purchase

The British Colonies in 1770

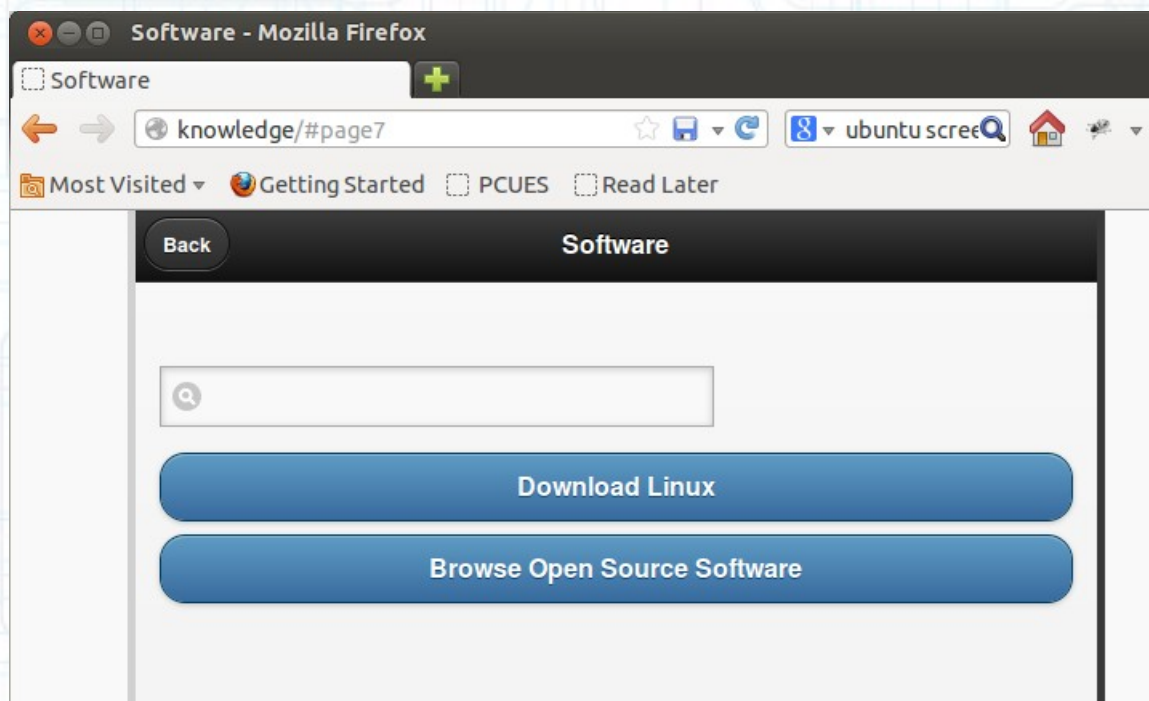
U.S. TERRITORIAL ACQUISITIONS

Called to Great Britain 1818

CANADA



# Software



# Made Possible By

- **Books** – Project Gutenberg
  - **Joel Steres** (book browser)
- **Wikipedia** - Kiwix and OpenZim
  - **Emmanuel Engelhart** (kiwix)
- **Hardware** – Arch Arm
  - **Mike Brown** and **Jeff Doozan** (uboot)
- **SGVHAK** – James, Michael S, Doug, et al
- **Maps** – Open Street Map
- **Video** – Khan Academy, Zurd torrent
- **Software** – Ubuntu, Linux community



# Resources

**Braddock Gaskill**

Internet-in-a-Box Project Lead

[braddock@braddock.com](mailto:braddock@braddock.com)

<http://internet-in-a-box.org>

# Try It Now!

## **Wifi Hotspot:**

**SSID: internet in a box**

Once connected, go to any web page

**Or ("wired" version)**

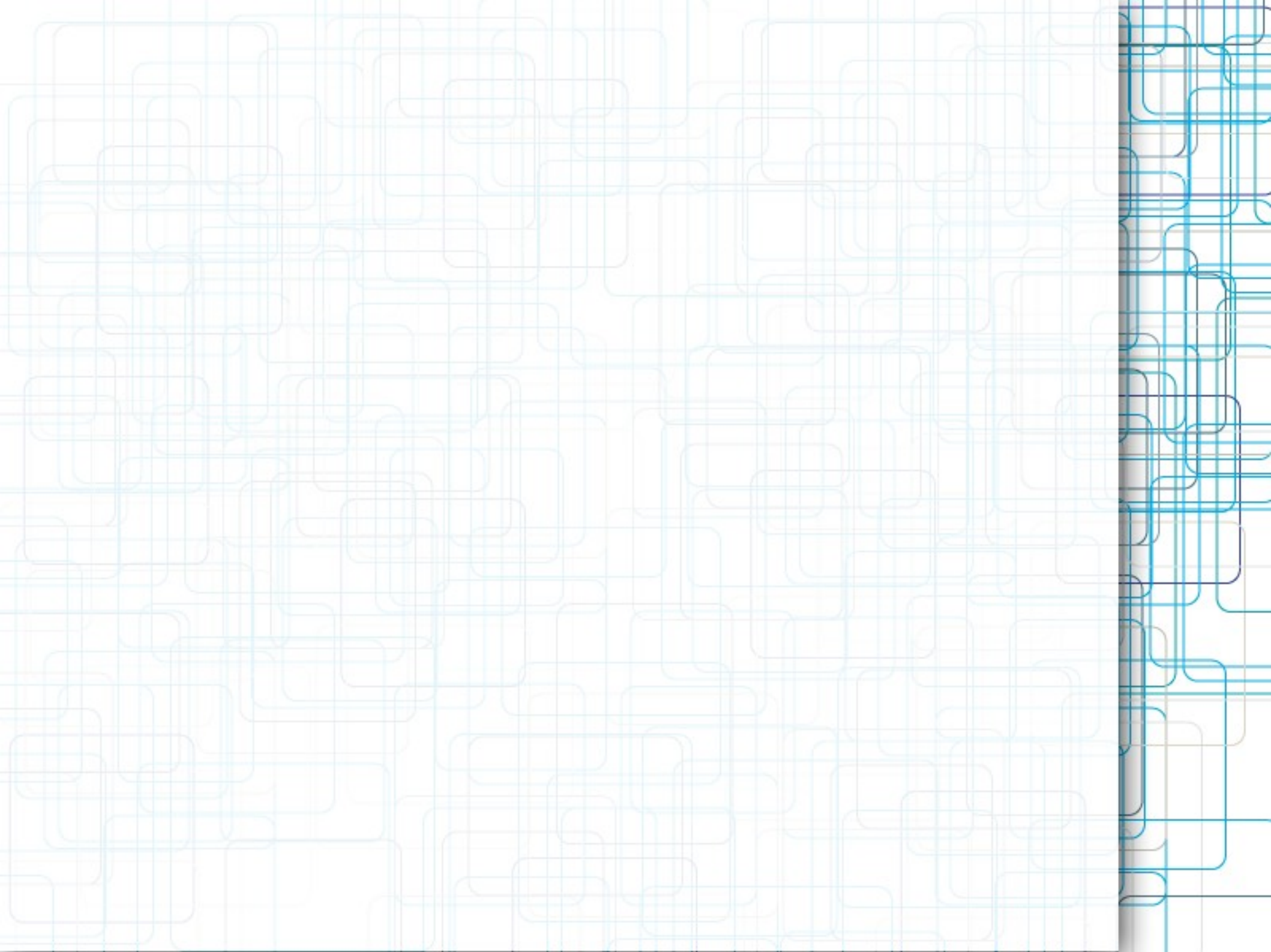
**SSID: internet in a box wired**

**<http://knowledge>**

or

**<http://192.168.1.113>**





# TECHNICAL PRESENTATION

(Mis)-Adventures in Data Processing



# Meet the Hardware

## Seagate Wireless Plus

- \$200
- Wireless Harddrive
- 1TB
- Argo Linux firmware
- Telnet root/goflex
- Uboot – working on Debian
- We run chroot debian



# Meet the Hardware

## Seagate GoFlex Home

- 1TB Low-end NAS
- \$150
- Ethernet 1Gbps
- Arch Arm Linux installed
- Can run Debian
- We run chroot Debian in Arch (don't ask!)





# Meet the Hardware

## Behind the scenes

- Job Server "**Node-8**"
  - 8GB RAM, Quad Core 2.5GHz, 1TB HDD
- Job Server "**Node-16-SSD**"
  - **16GB RAM, 500GB SSD**, Quad Core 3GHz, 6TB HDD

# Maps

- Open Street Maps
  - Open Source effort to map the world
  - Merge data from many sources (government, donated, created)
  - Open Data Commons Open Database License (ODbL)
  - LeafletJS JavaScript app



# Maps Process

1. Download "Planet" File (**30GB** binary)
2. Import Planet into PostgreSQL DB
  - **350 GB** final DB size (!)
  - Import took 2 weeks (!) on Node-8, 4 days on Node-16-SSD
3. Render DB to image tiles (PNG)
  - 1 week with Node-8 & Node-16 in cluster
  - **~150 GB** of image tiles
4. Pack image tiles
  - **~100 GB**

# Serving Maps

- The GoFlex and Wireless Plus are too slow to render maps
  - I know, I built the entire stack and tried it – 1 to 5 minutes per tile
- We pre-render **EVERYTHING**
- 4,096 PNG images per file
- Redundant tiles removed (many many!)
- Python back-end extracts and serves images



# Books

- Project Gutenberg, founded 1971
- 40,000 e-books, mostly Public Domain
- Custom Web UI (Thanks Joel!)

# Books Process

- Project Gutenberg Archive
  - PG Archive is a mess
  - Many formats, most redundant
  - Ex: archive contains ISOs of prior versions of the archive (!)
- Mirror
  - Rsync from ibiblio (service to be discontinued)
  - Rsync service slow (throttled?) and unreliable
  - **~1.5 weeks**



# Wikipedia

- Seventh highest ranked web site in the world
- Fully Open Source content

# Wikipedia

- Using OpenZim and Kiwix
- OpenZim
  - A container file format designed FOR Wikipedia
  - Wikipedia pre-rendered to HTML, HTML stored in Zim file
- Kiwix
  - Offline Wikipedia reader based on Zim
  - Desktop reader and Zim server
  - We reverse proxy kiwix-serve (for now)



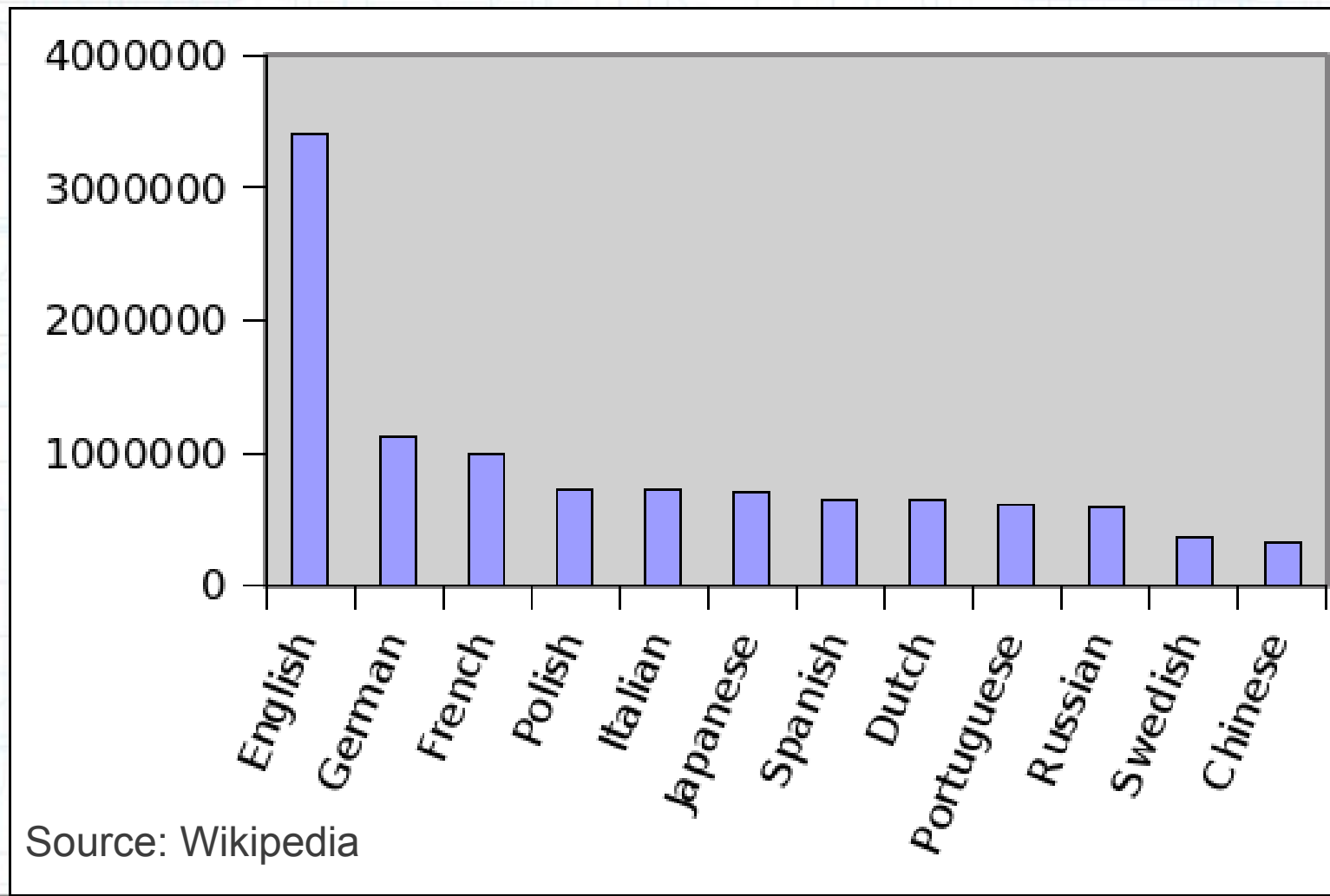
# Wikipedia: The Data

- Direct Download
  - All Wikimedia projects have XML dump files available for download
  - No bulk download option for images
- Zim community
  - Provides pre-rendered Wikipedia ZIM torrents for 40 languages
  - Many include scaled images
  - Some a few years stale

# Wikipedia

- Not rich in all languages

Number of articles for top languages





# Future: Open Source Textbooks

- Aka "Open Educational Resources" (OER)
- Many models emerging
  - State-funded Model
    - California Open Source Textbook Project
    - State of California spends \$400 million per year on textbooks
  - Volunteer-driven Model
    - South Africa – Free High School Science Texts (FHSST)
      - Math and Science high school curriculum
    - WikiBooks Project
      - 2,000+ books "in progress", few finished

# Future: Open Source Textbooks

- Many models emerging
  - Publisher Model
    - Flat World Knowledge (Publisher)
    - 40+ Textbooks, all English
    - Publisher and author make money off print copies and formatted downloads
  - University-driven Model
    - MIT Open Courseware
    - College course material provided both for students and others



# WorldReader Project

- Deploys commercial e-book readers



# Humane Reader





# Humane Reader

- Displays to television set
- Reads Micro-SD Card (2GB)
- Optional PS/2 Keyboard
- Wikipedia reader
- \$20 in quantity



# Humane Reader

- Minimalist design
  - Uses two 8-bit microcontrollers
  - Specs of a 1977 Apple II computer
- Cell phone standard micro-USB power
- Open Source Design
  - Effort to build community around platform
- Arduino expansion slots
- Can be used as simple 8-bit computer



# Resources

**Braddock Gaskill**

Internet-in-a-Box Project Lead

[braddock@braddock.com](mailto:braddock@braddock.com)

<http://internet-in-a-box.org>