Can science be wikified?
(and should it)

Ivar Martin
Knowen.org
Outline

• Does wikification makes sense?
Outline

- Does wikification makes sense?
- How do we get there?
Wikipedia

Everyone can read and edit

- 18B page views and 500M unique visitors/month (Alexa rank 5)
- Number of articles: 5.5M (English only), 40K “high quality”
- 33M registered editors, 140K active editors
- 50% edits are done by 500 people, 0.7% (2009)

A failure of science online: Wikipedia
Wikipedia is a second example where scientists have missed an opportunity to innovate online. Wikipedia has a vision statement to warm a scientist’s heart: “Imagine a world in which every single human being can freely share in the sum of all knowledge. That’s our commitment.”

Nowadays, Wikipedia’s success has to some extent legitimized contribution within the scientific community. But how strange that the modern day Library of Alexandria had to come from outside academia.
Wikipedia appeal

- Universal – central – repository
- Generally good high-level articles
- Active participation (billions of users, thousands of contributors)
- As close as can get to
If Wikipedia were used for science

• Potential benefits
  – Universal resource, instead of zillions of journals
  – Existing context for new results
  – Fairer attribution

• Potential problems
  – Public quarrels
  – Fractionalized content ("hard forks")
  – ...
Why scientist don’t use Wikipedia
Why scientist don’t use Wikipedia

- Mixed quality of material
Why scientist don’t use Wikipedia

- Mixed quality of material
- No tangible credit to contributor
Why scientist *don’t* use Wikipedia

• Mixed *quality* of material
• No tangible *credit* to contributor
• Not a place for *original* material
Why scientist *don’t* use Wikipedia

- Mixed *quality* of material
- No tangible *credit* to contributor
- Not a place for *original* material
- *Transience* (editing over)
Why scientist don’t use Wikipedia

• Mixed quality of material
• No tangible credit to contributor
• Not a place for original material
• Transience (editing over)
• Consensus: not (always) expertise based
Why scientist don’t use Wikipedia

- Mixed quality of material
- No tangible credit to contributor
- Not a place for original material
- Transience (editing over)
- Consensus: not (always) expertise based
- Content intended for non-experts
Wikipedia-for-science requirements
Wikipedia-for-science requirements

- **Incentives**
  - Improve *visibility* of work
  - Enhance personal *productivity* and *quality* of work
    - Note taking, long-term preservation
  - Improve *collaborations*
  - Build *reputation*
    - Track/value all contributions
Wikipedia-for-science requirements

- **Incentives**
  - Improve visibility of work
  - Enhance personal **productivity** and **quality** of work
    - Note taking, long-term preservation
  - Improve **collaborations**
  - Build **reputation**
    - Track/value all contributions

- **Access control**
  - Public or Limited access
Wikipedia-for-science requirements

- **Incentives**
  - Improve *visibility* of work
  - Enhance personal *productivity* and *quality* of work
    - Note taking, long-term preservation
  - Improve *collaborations*
  - Build *reputation*
    - Track/value all contributions

- **Access control**
  - Public or Limited access

- **Structure**
  - Levels of *refinement*
Wikipedia-for-science requirements

- **Incentives**
  - Improve *visibility* of work
  - Enhance personal *productivity* and *quality* of work
    - Note taking, long-term preservation
  - Improve *collaborations*
  - Build *reputation*
    - Track/value all contributions
- **Access control**
  - Public or Limited access
- **Structure**
  - Levels of *refinement*
- **AI**
  - Suggest connections to existing content
  - Analytics tools
    - performance,
    - trends
Incentives

• StackExchange model
  – Reputation, badges, karma
Incentives

- StackExchange model
  - Reputation, badges, karma

- Blockchain model
  - Tokens (utility, work), convertible into $
Incentives

- StackExchange model

Xiao-Gang Wen [top 20% overall]

I am a professor working on condensed matter theory. My current interest is in topological order, which correspond to patterns of long-range entanglement in many-body system.

Understanding patterns of many-body entanglement is related to some modern mathematics. For example, the math framework for 2D long-range entanglements happen to be unitary fusion category theory. For higher dimensions, we may need higher categories. The math framework for short-range entanglements with symmetry happen to be group cohomology theory of the symmetry group and cobordism theory.
Incentives

- StackExchange model

Terry Tao
Professor of Mathematics at UCLA

[Image of StackExchange website]
Structure
What is the structure of knowledge?
What is the structure of knowledge?

• Collections of facts -> connection of facts
  – Generalization/refinement
  – Efficient Compression (sparse recovery/compressed sensing)
What is the structure of knowledge?

- **Collections of facts -> connection of facts**
Knowledge hierarchies

- **Math**
  - Mathematical Subject Classification – MSC2010: https://zbmath.org/classification/
  - Wolfram: http://mathworld.wolfram.com/
  - NLab: https://ncatlab.org/nlab/show/mathematics
  - Enc of Math: https://www.encyclopediaofmath.org/index.php/Talk:EnM:This_project#Categories

- **Physics**
  - https://physh.aps.org/about
  - PhySH: https://physh.aps.org/browse
  - PACS: https://www.aip.org/publishing/pacs/pacs-2010-regular-edition
  - NLab: https://ncatlab.org/nlab/show/physics

- **Medicine**

- **Computer science**
  - Computing classification system: http://dl.acm.org/ccs/ccs.cfm
  - Computing research repository: http://arxiv.org/corr/subjectclasses

- **Economics**

- **General**
  - Lib of Congress: https://www.loc.gov/catdir/cpso/lcco/
Is knowledge a tree?

Science

Physics

Biology

Mechanics

Electricity & Magnetism

Quantum Field Theory

Specialization
Generalization
Analogy
Is knowledge a tree?

Can it be shaped into a tree?
Inspiration from RG/Machine learning
Inspiration from RG/Machine learning

Statistics > Machine Learning

An exact mapping between the Variational Renormalization Group and Deep Learning

Pankaj Mehta, David J. Schwab

(Submitted on 14 Oct 2014)
# Inspiration

THE ART OF COMPUTER PROGRAMMING
THIRD EDITION

## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 1 — Basic Concepts</td>
<td>1</td>
</tr>
<tr>
<td>1.1. Algorithms</td>
<td>1</td>
</tr>
<tr>
<td>1.2. Mathematical Preliminaries</td>
<td>10</td>
</tr>
<tr>
<td>1.2.1. Mathematical Induction</td>
<td>11</td>
</tr>
<tr>
<td>1.2.2. Numbers, Powers, and Logarithms</td>
<td>27</td>
</tr>
<tr>
<td>1.2.3. Sum and Products</td>
<td>59</td>
</tr>
<tr>
<td>1.2.4. Integer Functions and Elementary Number Theory</td>
<td>45</td>
</tr>
<tr>
<td>1.2.5. Permutations and Factorials</td>
<td>52</td>
</tr>
<tr>
<td>1.2.6. Binomial Coefficients</td>
<td>75</td>
</tr>
<tr>
<td>1.2.7. Harmonic Numbers</td>
<td>79</td>
</tr>
<tr>
<td>1.2.8. Fibonacci Numbers</td>
<td>87</td>
</tr>
<tr>
<td>1.2.9. Generating Functions</td>
<td>96</td>
</tr>
<tr>
<td>1.2.11. Asymptotic Representations</td>
<td>107</td>
</tr>
<tr>
<td>1.2.11.1. The Olnomical</td>
<td>107</td>
</tr>
<tr>
<td>1.2.11.2 Euler's summation formula</td>
<td>111</td>
</tr>
<tr>
<td>1.2.11.3 Some asymptotic calculations</td>
<td>114</td>
</tr>
<tr>
<td>1.3. MIX</td>
<td>124</td>
</tr>
<tr>
<td>1.3.1. Description of MIX</td>
<td>134</td>
</tr>
<tr>
<td>1.3.2. The MIX Assembly Language</td>
<td>144</td>
</tr>
<tr>
<td>1.3.3. Applications to Permutations</td>
<td>164</td>
</tr>
<tr>
<td>1.4. Some Fundamental Programming Techniques</td>
<td>196</td>
</tr>
<tr>
<td>1.4.1. Subroutines</td>
<td>198</td>
</tr>
<tr>
<td>1.4.2. Closures</td>
<td>200</td>
</tr>
<tr>
<td>1.4.3. Interpretive Routines</td>
<td>202</td>
</tr>
<tr>
<td>1.4.3.1. A MIX simulator</td>
<td>212</td>
</tr>
<tr>
<td>1.4.4. Input and Output</td>
<td>215</td>
</tr>
<tr>
<td>1.4.5. History and Bibliography</td>
<td>229</td>
</tr>
<tr>
<td>Chapter 2 — Information Structures</td>
<td>232</td>
</tr>
<tr>
<td>2.1. Introduction</td>
<td>232</td>
</tr>
<tr>
<td>2.2. Linear Lists</td>
<td>233</td>
</tr>
<tr>
<td>2.2.1. Stacks, Queues, and Deques</td>
<td>233</td>
</tr>
<tr>
<td>2.2.3. Sequential Allocation</td>
<td>244</td>
</tr>
<tr>
<td>2.2.3. Linked Allocation</td>
<td>254</td>
</tr>
<tr>
<td>2.2.4. Circular Lists</td>
<td>273</td>
</tr>
<tr>
<td>2.2.5. Doubly Linked Lists</td>
<td>290</td>
</tr>
<tr>
<td>2.2.6. Arrays and Orthogonal Lists</td>
<td>298</td>
</tr>
<tr>
<td>2.3. Trees</td>
<td>336</td>
</tr>
<tr>
<td>2.3.1. Traversing Binary Trees</td>
<td>318</td>
</tr>
<tr>
<td>2.3.2. Binary Tree Representation of Trees</td>
<td>334</td>
</tr>
<tr>
<td>2.3.3. Other Representation of Trees</td>
<td>340</td>
</tr>
<tr>
<td>2.3.4. Basic Mathematical Properties of Trees</td>
<td>362</td>
</tr>
<tr>
<td>2.3.4.1. Free trees</td>
<td>363</td>
</tr>
<tr>
<td>2.3.4.2. Oriented trees</td>
<td>372</td>
</tr>
<tr>
<td>2.3.4.3. The &quot;infinity lemma&quot;</td>
<td>382</td>
</tr>
<tr>
<td>2.3.4.4. Enumeration of trees</td>
<td>386</td>
</tr>
<tr>
<td>2.3.4.5. Path length</td>
<td>399</td>
</tr>
<tr>
<td>2.3.4.6. History and bibliography</td>
<td>406</td>
</tr>
<tr>
<td>2.3.5. Lists and Garbage Collection</td>
<td>409</td>
</tr>
<tr>
<td>2.4. Multilinked Structures</td>
<td>424</td>
</tr>
<tr>
<td>2.5. Dynamic Storage Allocation</td>
<td>435</td>
</tr>
<tr>
<td>2.6. History and Bibliography</td>
<td>457</td>
</tr>
</tbody>
</table>

## Answers to Exercises

Appendix A — Tables of Numerical Quantities

1. Fundamental Constants (decimal)                  | 619   |
2. Fundamental Constants (octal)                    | 619   |
3. Harmonic Numbers, Bernoulli Numbers, Fibonacci Numbers | 621   |

Appendix B — Index to Notations

Index and Glossary                                  | 628   |
Why Tree or DAG structure is good

- depth is $\sim \log N$
- Easy to navigate
- Easy to learn
- “Plug-and-Play”
How to get there?
How to get there?

Chicken or Egg?
How to get there?

• Seed from existing content
  – Wikipedia
  – Published articles
  – Taxonomies/classification schemes
How to get there?

• Seed from existing content
  – Wikipedia
  – Published articles
  – Taxonomies/classification schemes

• Organically, new collaborative projects
  – Bottom up
  – Knowen
Incommensurate frequency Floquet

Main idea:

Driving a system with multiple incommensurate frequencies leads to Floquet lattices in multiple dimensions. Various lattice Hamiltonians can be implemented, including topologically non-trivial ones. The number of bands can be controlled by choosing appropriate building blocks -- two-level systems will lead to two-band Hamiltonians, 3-level → 3 band, etc.

The distinguishing feature of such Floquet lattice Hamiltonians is that they are tilted, with the tilt potential $U(n_1, n_2, \ldots) = n_1 \omega_1 + n_2 \omega_2 + \ldots$. That has several consequences:

- The wave-functions are localized in the direction perpendicular to the hyperplane $U(n_1, n_2, \ldots) = \text{const}$.
  - for weak drive the wavefunction will be localized on the lattice points closest to the hyperplanes.
  - The lattice potential at these points is quasi-periodic.
  - for strong drive, w.f. will be weakly confined, as if they are near a smooth edge of a sample.

- if band structure is topologically non-trivial, every hyperplane is also a locus of topologically protected "edge-modes." For instance, in the case of two-dim Floquet space, there will be helical modes propagating along the lines $n_1 \omega_1 + n_2 \omega_2 = \text{const}$. Chiral propagation correspond to preferential absorption of one frequency photon and emission of the other.

- Even when the drive frequencies are commensurate, the pumping between the modes persists as can be
Elements

• Nodes: (Text, data, scripts)
• Commits and commenting (cf. github)
• Structure: Tree/Directed acyclic graph (DAG)
• Access: public and private projects
• Reputation tracking

Also: collaborative editing, subscriptions, feedback mechanisms
Consider a zero-dimensional system with basis states \( \alpha \) subject to time-dependent Hamiltonian \( H_{\alpha}(t) \).

\[
\dot{\psi}_\alpha(t) = H_{\alpha}(t)\psi_\alpha(t).
\]  

(1)

We will assume that \( H(t) = H(\omega_1 t, \omega_2 t) \), with \( H(\theta_1, \theta_2) \) being \( 2\pi \)-periodic in \( \theta_1, \theta_2 \). Let us now introduce the Floquet representation for the wave function (appropriate for multi-tone drive)

\[
\psi(t) = e^{-iE_\alpha t}\phi(t) = e^{-iE_{\alpha\omega_1 \omega_2} t + i\omega_1 \alpha_1 + i\omega_2 \alpha_2} \phi_{\alpha_1 \alpha_2}
\]  

(2)

and expand Hamiltonian in terms of its Fourier components with respect to \( \theta_1, \theta_2 \),

\[
H(\theta_1 + \omega_1 t, \theta_2 + \omega_2 t) = \sum_{\alpha_1 \alpha_2} \hat{h}_{\alpha_1 \alpha_2} e^{-i\omega_1 \alpha_1 - i\omega_2 \alpha_2}.
\]

Now, substituting into the Schrödinger equation Eq. (1), we get,

\[
(E + \omega_1 \alpha_1 + \omega_2 \alpha_2)\phi_{\alpha_1 \alpha_2} = \hat{h}_{\alpha_1 \alpha_2} \phi_{\alpha_1 \alpha_2}
\]  

(3)

This is the key equation that describes hopping in the Floquet space. It can be solved by diagonalizing a sufficiently large block of it. The integer coordinates
Public-Private Division

- Public

- Private
Collaboration tool – most projects are **private**

**Workflow:**
- Formulate Idea – *create top level node*
- Explore sub-branches of idea/sub-ideas – *create subnodes*
- get results/refine idea – *go back and forth between nodes*

**All along:**
- Invite collaborators
- Track and comment on any changes
- Build reputation
Public-Private Fusion
(“plug-and-play”)

• Public

• Private
Public-Private Fusion ("plug-and-play")

- Public
- Private
- Private Visible
Public-Private Fusion
(“plug-and-play”)

- Public
- Private
- Private Visible
- Public
Public-Private Fusion
(“plug-and-play”)
Vision

- Global (decentralized) structured wiki-like science resource (knowen.org is an experimental prototype)
- Uses:
  - Outreach
  - Learning
  - Sharing results/methods/data
  - Raw material for packaging into books and reviews
- Reputation/priority recording
- Intelligent Automatic suggestions for content placement and search

Can exist in parallel with current article/journal infrastructure
Sustained by community (cf blockchain), or nonprofits (cf arxiv)