Computational Thinking in the Sciences and Beyond

Jeannette M. Wing
Corporative Vice President, Microsoft Research

Microsoft Research
Asia Faculty Summit – Beijing China
30 October 2014
My Grand Vision

• **Computational thinking** will be a fundamental skill used by everyone in the world by the middle of the 21st Century.

  – Just like reading, writing, and arithmetic.
  – Incestuous: Computing and computers will enable the spread of computational thinking.

  – In **research**: scientists, engineers, ..., historians, artists
  – In **education**: K-12 students and teachers, undergrads, ...

What is Computational Thinking?

Computational thinking is the *thought processes* involved in formulating a problem and expressing its solution(s) in such a way that a computer—human or machine—can effectively carry out.
Computational Thinking, Philosophically

• Complements and combines mathematical and engineering thinking
  – C.T. draws on math as its foundations
    • But we are constrained by the physics of the underlying machine
  – C.T. draws on engineering since our systems interact with the real world
    • But we can build virtual worlds unconstrained by physical reality

• Ideas, not artifacts
  – It’s not just the software and hardware that touch our daily lives, it will be the computational concepts we use to approach living.

• It’s for everyone, everywhere
Sample Classes of Computational Concepts

- **Algorithms**
  - E.g., mergesort, binary search, string matching, clustering
- **Data Structures**
  - E.g., sequences, tables, trees, graphs, networks
- **State Machines**
  - E.g., finite automata, Turing machines
- **Languages**
  - E.g., regular expressions, ..., VDM, Z, ..., ML, Haskell, ..., Java, Python
- **Logics and semantics**
  - E.g., Hoare triples, temporal logic, modal logics, lambda calculus
- **Heuristics**
  - E.g., A* (best-first graph search), caching
- **Control Structures**
  - Parallel/sequential composition, iteration, recursion
- **Communication**
  - E.g., synchronous/asynchronous, broadcast/P2P, RPC, shared memory/message-passing
- **Architectures**
  - E.g., layered, hierarchical, pipeline, blackboard, feedback loop, client-server, parallel, distributed, fault-tolerant

**NOT**

- Computer literacy, i.e., how to use Word and Excel or even Google or Bing
- Computer programming, i.e., beyond Java Programming 101
Examples of Computational Thinking in Other Disciplines
One Discipline, Many Computational Methods
Computational Thinking in Biology

- Shotgun algorithm expedites sequencing of human genome
- Abstract interpretation in systems biology
- Model checking applied to arrhythmia, diabetes, pancreatic cancer
- DNA sequences are strings in a language
- Boolean networks approximate dynamics of biological networks
- Cells as a self-regulatory system are like electronic circuits
- Process calculi model interactions among molecules
- Statecharts used in developmental genetics
- Protein kinetics can be modeled as computational processes
- Robot Adam discovers role of 12 genes in yeast
- PageRank algorithm inspires ecological food web

Insight: Models and languages for expressing computational processes are good for expressing the dynamics of biological processes.
Stem Cell Prediction

Embryonic Stem (ES) Cells

Self-renewing: Divide indefinitely

Pluripotent: Generate all adult cell types, and can be re-injected back into the developing embryo

We can reprogram adult cells to this state

Transient in tissue: A culture-dependent phenomenon
Extrinsic Signals Control ES Cell Behaviour

Whether an ES cell will remain self-renewing, or differentiate towards an adult cell lineage depends on the signals that it receives.

The signals required to sustain ES cells in culture have been progressively refined. Any two of LIF, CH and PD are sufficient.

We can measure the expression of key genes under different combinations of signals to gain insight into the dynamic behavior of the system.
A Reasoning Engine for Interaction Networks

RE:IN is a tool built on Z3 that utilizes Satisfiability Modulo Theories to synthesize only those (out of $10^{43}$) Boolean networks that provably satisfy experimental constraints.
The set of possible models was constrained by experimentally-observed behaviours.

This set was used to make a large number (53) of non-intuitive predictions of the response of the network to genetic perturbations. These predictions were experimentally validated with over 70% accuracy rate.

The highlighted interactions show the minimal set required to explain stem cell behaviour: **the essential program governing naïve pluripotency**
One Computational Method,
Many Disciplines

Machine Learning has transformed the field of Statistics.
Machine Learning in the Sciences

- Brown dwarfs and fossil galaxies discovery via machine learning, data mining, data federation
- Very large multi-dimensional datasets analysis using KD-trees

Astronomy

- Anti-inflammatory drugs
- Chronic hepatitis
- Mammograms
- Renal and respiratory failure

Medicine

- Tornado formation

Meteorology

- fMRI data analysis to understand language via machine learning

Neurosciences
Spammers mutate their messages to work around filters

Solution: Go after the weak link

HIV mutates to avoid attack by immune system

Solution: Go after the weak link
Strategy: Identify vulnerable regions on HIV and create a vaccine that directs the immune system to target those regions.
Finding Vulnerable Regions

- Identify a set of “controllers” who are infected with HIV, but don’t get very sick
- Look for differences between where immune systems of controllers vs where those of normal people are attacking HIV
- These differences point to the vulnerable regions

<table>
<thead>
<tr>
<th>amino acid sequence</th>
<th>protein</th>
<th>begin</th>
<th>end</th>
</tr>
</thead>
<tbody>
<tr>
<td>KAFSPEVIPMF</td>
<td>p24</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>RLRDLLLIVTR</td>
<td>gp41</td>
<td>259</td>
<td>269</td>
</tr>
<tr>
<td>GIPHPAGLK</td>
<td>pr</td>
<td>192</td>
<td>200</td>
</tr>
<tr>
<td>HTQGYFPDW</td>
<td>nef</td>
<td>116</td>
<td>124</td>
</tr>
<tr>
<td>AEAMSQVTNS</td>
<td>p2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>SAEPVPLQL</td>
<td>rev</td>
<td>67</td>
<td>75</td>
</tr>
<tr>
<td>QAISPRTLNAW</td>
<td>p24</td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>RIKQINMW</td>
<td>gp120</td>
<td>419</td>
<td>427</td>
</tr>
</tbody>
</table>

From data from hundreds of controllers and non-controllers, machine learning helped identified 8 regions of HIV proteins where controllers are much more likely to attack.
Finding More Vulnerable Regions

Use in silico prediction under the assumption that vulnerable regions are where the amino acid changes protein stability (increases “structural entropy”).

Estimate protein stability with simulations using the FoldX algorithm.

These regions are new candidate vulnerable regions.
Machine Learning Everywhere

- Credit Cards
- Supermarkets
- Wall Street
- Entertainment: Shopping, Music, Travel
- Sports
- Netflix
- Amazon Prime
- Travelocity
Computational Thinking in the Sciences and Beyond
Computational Thinking in Other Sciences

**Chemistry**
- Atomistic calculations are used to explore chemical phenomena
- Optimization and searching algorithms identify best chemicals for improving reaction conditions to improve yields

**Physics**
- Adiabatic quantum computing: How quickly is convergence?
- Genetic algorithms discover laws of physics.

**Geosciences**
- Abstractions for Sky, Sea, Ice, Land, Life, People, etc.
- Hierarchical, composable, modular, traceability, allowing multiple projections along any dimension, data element, or query
- Cornell’s NSF Expedition on Computational Sustainability
Computational Thinking in Math and Engineering

**Mathematics**

- Discovering E8 Lie Group: 18 mathematicians, 4 years and 77 hours of supercomputer time (200 billion numbers). Profound implications for physics (string theory)
- Four-color theorem proof

**Engineering (electrical, civil, mechanical, aero & astro,...)**

- Calculating higher order terms implies more precision, which implies reducing weight, waste, costs in fabrication
- Boeing 777 tested via computer simulation alone, not in a wind tunnel
- Hybrid automata for modeling and analyzing cyber-physical systems

Credit: Wikipedia
Credit: Boeing
Computational Thinking for Society

**Economics**
- Automated mechanism design underlies electronic commerce, e.g., ad placement, on-line auctions, kidney exchange
- Internet marketplace requires revisiting Nash equilibria model
- Use intractability for voting schemes to circumvent impossibility results

**Law**
- Inventions discovered through automated search are patentable
- Stanford CL approaches include AI, temporal logic, state machines, process algebras, Petri nets
- POIROT Project on fraud investigation is creating a detailed ontology of European law
- Sherlock Project on crime scene investigation

**Healthcare**
- Algorithmic medicine
- Software design principles and debugging applied to prescriptions of painkillers
- ONC SHARP Program, NSF Smart Health and Wellness Program, NITRD Senior Steering Group on Health IT
Fast and Accurate Decision-Making

2012 Xbox

• Polling daily and during live debates
• 350k respondents: age, race, gender, state, edu, party id, ideology, previous votes, registration
• 750k polls in 45 days
• 30,000 respondents 5 or more polls
• Data and engagement great, but accuracy?

Cite: Forecasting elections with non-representative polls; Gelman, Goel, Rothschild, and Wang (2014)
Data Collection

Two-party Obama Support

Cite: Forecasting elections with non-representative polls; Gelman, Goel, Rothschild, and Wang (2014)
Data Analytics

Use convex optimization to keep millions of interrelated predictions consistent

Two-party Obama Support

Cite: Forecasting elections with non-representative polls; Gelman, Goel, Rothschild, and Wang (2014)
Fifa World Cup 2014: Microsoft beats Google and Siri in predictions

By Joseph George
Published Tuesday, July 15, 2014

It was Paul the Octopus in 2010, but this time around, it is Microsoft's Cortana that has emerged winner in predicting World Cup results.

Cortana, powered by Bing, got almost all the results correct except for one, thereby overtaking both Google and Apple’s Siri.

"Thanks to Bing, Cortana has accurately predicted the winner in 15 out of 16 games," said the latest blog post by Windows.

Bing launched its Bing Predicts ahead of the world Cup and got 100 percent matches and final elimination -it picked Brazil as
Computational Thinking for Society

Archeology
- eHeritage Project, Microsoft Research Asia
- Digital Forma Urbis Romae Project, Stanford
- Cathedral Saint Pierre, Columbia

- Crowd sourcing as a new way of getting news tips from sources
- Algorithmic approach to validate credibility of sources
- Digital Media and Learning Initiative, MacArthur Foundation

Journalism

Humanities
- Digging into Data Challenge: What could you do with a million books?
  Nat’l Endowment for the Humanities (US), JISC (UK), SSHRC (Canada)
- Music, English, Art, Design, Photography, ...
Computational Social Science: Learning about Crowdworkers

- **Computational**: digital studies produce the *nodes* (people) and *edges* (relationships) in a network

- **Anthropology**: qualitative studies produce the *variety of nodes* (individuals, institutions) and *meaning of edges* (motivations, hierarchies, power dynamics)
Mapping the Crowd

Self-reported locations for ~10,000 participants in a map task on Amazon Mechanical Turk. Coloration of counties/districts is by population density.
Computational Thinking in Education
Pre-K to Grey

K-6, 7-9, 10-12

• Undergraduate courses
  – Freshmen year
    • “Ways to Think Like a Computer Scientist” aka Principles of Computing
  – Upper-level courses

• Graduate-level courses
  – Computational arts and sciences
    • E.g., entertainment technology, computational linguistics, …, computational finance, …, computational biology, computational astrophysics

• Post-graduate
  – Executive and continuing education, senior citizens
  – Teachers, not just students
Education Implications for K-12

Question and Challenge for the Computing Community:

What is an effective way of learning (teaching) computational thinking by (to) K-12?

- What concepts can students (educators) best learn (teach) when?  
  What is our analogy to numbers in K, algebra in 7, and calculus in 12?

- We uniquely also should ask how best to integrate The Computer with teaching the concepts.

Computer scientists are now working with educators and cognitive learning scientists to address these questions.
Computational Thinking in Daily Life
Getting Morning Coffee at the Cafeteria

- coffee
- soda
- straws, stirrers, milk
- cups
- sugar, creamers
- napkins
- lids
Getting Morning Coffee at the Cafeteria

- coffee
- soda
- straws, stirrers, milk
- cups
- sugar, creamers
- napkins
- lids
Getting Morning Coffee at the Cafeteria

Especially Inefficient With Two or More Persons...

- coffee
- soda
- straws, stirrers, milk
- cups
- sugar, creamers
- lids
- napkins
Better: Think Computationally—Pipelining!

- coffee
- soda
- straws, stirrers, milk
- cups
- sugar, creamers
- napkins
- lids
United States Efforts

High School

CS Principles: [http://csprinciples.org](http://csprinciples.org)
- With NSF support, revision of CS AP courses

K-12

[http://www.csta.acm.org/](http://www.csta.acm.org/)
- Computational Thinking Resource Set: A Problem-Solving Tool for Every Classroom
- K-12 Computer Science Standards

CSTB Reports:
- The Report of a Workshop on Pedagogical Aspects of Computational Thinking 2011
- Report of a Workshop on the Scope and Nature of Computational Thinking 2010

Congress

- proposed by PA Senator Casey and CO Representative Polis.
``“Computational thinking” offers insightful ways to view how information operates in many natural and engineered systems.

3. *Every child should have the opportunity to learn Computing at school.* We believe that:

- Every child should be expected to be ‘digitally literate’ by the end of compulsory education, in the same way that every child is expected to be able to read and write. “
Establish computer science as a foundational subject discipline, like math or physics, that every child should learn, from primary school onwards.

International Efforts

**Asia**

Computer Science (CS) Reloaded Programme

“...aims to deliver enrichment courses to pre-tertiary students to deepen their infocomm skills by supporting course fees for students to take up computer science courses anchored in computational thinking.”

**Europe**

- Computational Thinking
  - IST Austria
  - Heidelberg Institute for Theoretical Studies
  - HITS

**Latin America**

**Middle East**

- ParisTech
- University of Zurich
- CYTED

**Computational thinking and computer fundamental education**

Towards spreading computational thinking in Egypt and the Arab world.
Prof. Guoliang Chen (陈国梁院士) is one of the key influencers who put Computational Thinking as the core ability for CS Fundamental Courses in all China universities.

Prof. Guoliang Chen also started the first CS Fundamental Course of Computational Thinking.
In 2012, The Chinese Ministry of Education (MOE) announced a program to reform Computing Fundamental Courses focusing on Computational Thinking.

In 2014, MOE and Microsoft jointly put forward a new program to sponsor 16 full Computational Thinking Courses and 100 case studies.

Empower 7 Million students with Computational Thinking ability through CS Fundamental Courses.
La pensée informatique

par Jeannette M. Wing

Cet article fait suite aux divers interviews que nous avons faits et qui nous invitaient à une réflexion sur les fondements de notre discipline et ses aspects philosophiques et épistémologiques. Aujourd'hui l'article de Jeannette Wing nous conduit à réfléchir sur l'utilité et l'ubiquité de la pensée informatique et ses implications, mais aussi sur l'essence même de cette pensée.
Spread the Word

• Help make computational thinking commonplace!

To fellow faculty, students, researchers, administrators, teachers, parents, principals, guidance counselors, school boards, teachers’ unions, congressmen, policy makers, ...
Thank you!
References (Representative Only)

- **Computational Thinking**
- **Model Checking, Temporal Logic, Binary Decisions Diagrams**
- **Computational Thinking and Biology**
References (Representative Only)

- **Machine Learning and Applications**
  - Symbolic Aggregate Approximation, Eamonn Keogh, UC Riverside, [http://www.cs.ucr.edu/~eamonn/SAX.htm](http://www.cs.ucr.edu/~eamonn/SAX.htm) (applications in Medical, Meteorological and many other domains)
  - The Auton Lab, Artur Dubrawski, Jeff Schneider, Andrew Moore, Carnegie Mellon, [http://www.autonlab.org/autonweb/2.html](http://www.autonlab.org/autonweb/2.html) (applications in Astronomy, Finance, Forensics, Medical and many other domains)

- **Computational Thinking and Astronomy**
  - Sloan Digital Sky Survey @Johns Hopkins University, [http://www.sdss.jhu.edu/](http://www.sdss.jhu.edu/)

- **Computational Thinking and Archaeology**
  - See also Marc Levoy’s digital archaeology projects: [http://www-graphics.stanford.edu/~levoy/](http://www-graphics.stanford.edu/~levoy/)
  - See also UK universities: [http://en.wikipedia.org/wiki/Computational_archaeology#Research_groups_and_institutions](http://en.wikipedia.org/wiki/Computational_archaeology#Research_groups_and_institutions)

- **Computational Thinking and Chemistry**

- **Computational Thinking and Economics**
References (Representative Only)


• Computational Thinking and Journalism

• Computational Thinking and Law

• Computational Thinking and Medicine and Healthcare
  – Institute for Computational Medicine, Johns Hopkins University, http://www.icm.jhu.edu/
  – See also Symbolic Aggregate Approximation, Eamonn Keogh, UC Riverside, http://www.cs.ucr.edu/~eamonn/SAX.htm

• Computational Thinking and Meteorology
  – See also Symbolic Aggregate Approximation, Eamonn Keogh, UC Riverside, http://www.cs.ucr.edu/~eamonn/SAX.htm
References (Representative Only)

• Computational Thinking (especially Machine Learning) and Neuroscience

• Computational Thinking and Sports
  – Lance Armstrong’s cycling computer tracks man and machine statistics, website
Credits

- Copyrighted material used under Fair Use. If you are the copyright holder and believe your material has been used unfairly, or if you have any suggestions, feedback, or support, please contact: jsoleil@nsf.gov

- Except where otherwise indicated, permission is granted to copy, distribute, and/or modify all images in this document under the terms of the GNU Free Documentation license, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation license” (http://commons.wikimedia.org/wiki/Commons:GNU_Free_Documentation_License)