

From Galapagos to Twitter: Darwin, Natural Selection, and Web 2.0

Xavier Llorà

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
Urbana, Illinois, 61801

xllora@ncsa.illinois.edu

<http://www.ncsa.illinois.edu/~xllora>

What is this talk about?

- Simple answer:
Darwin's natural selection concept and information technologies
- Natural selection?
Natural selection is the process by which heritable traits that make it more likely for an organism to survive and successfully reproduce become more common in a population over successive generations.
- Remember, back on Darwin's time genetics were unknown



Darwin on HMS Beagle



A quick Darwin background check

- In Edinburgh Charles Darwin
 - Learnt standard geology (Robert Jameson)
 - Exposed to British catastrophism of 1820s
 - Run into Lamarckian ideas with Robert Grant
- Cambridge provide him a chance to
 - Under John Henslow, explored boundaries between varieties and species
 - Read and accepted traditional theological explanations of the world
- Evolution existed *before* and *after* Darwin
- HMS Beagle
 - Left Plymouth 2pm on December 27, 1831
 - Returned October 2, 1836



HMS Beagle



And Galapagos happened



Massive diversity



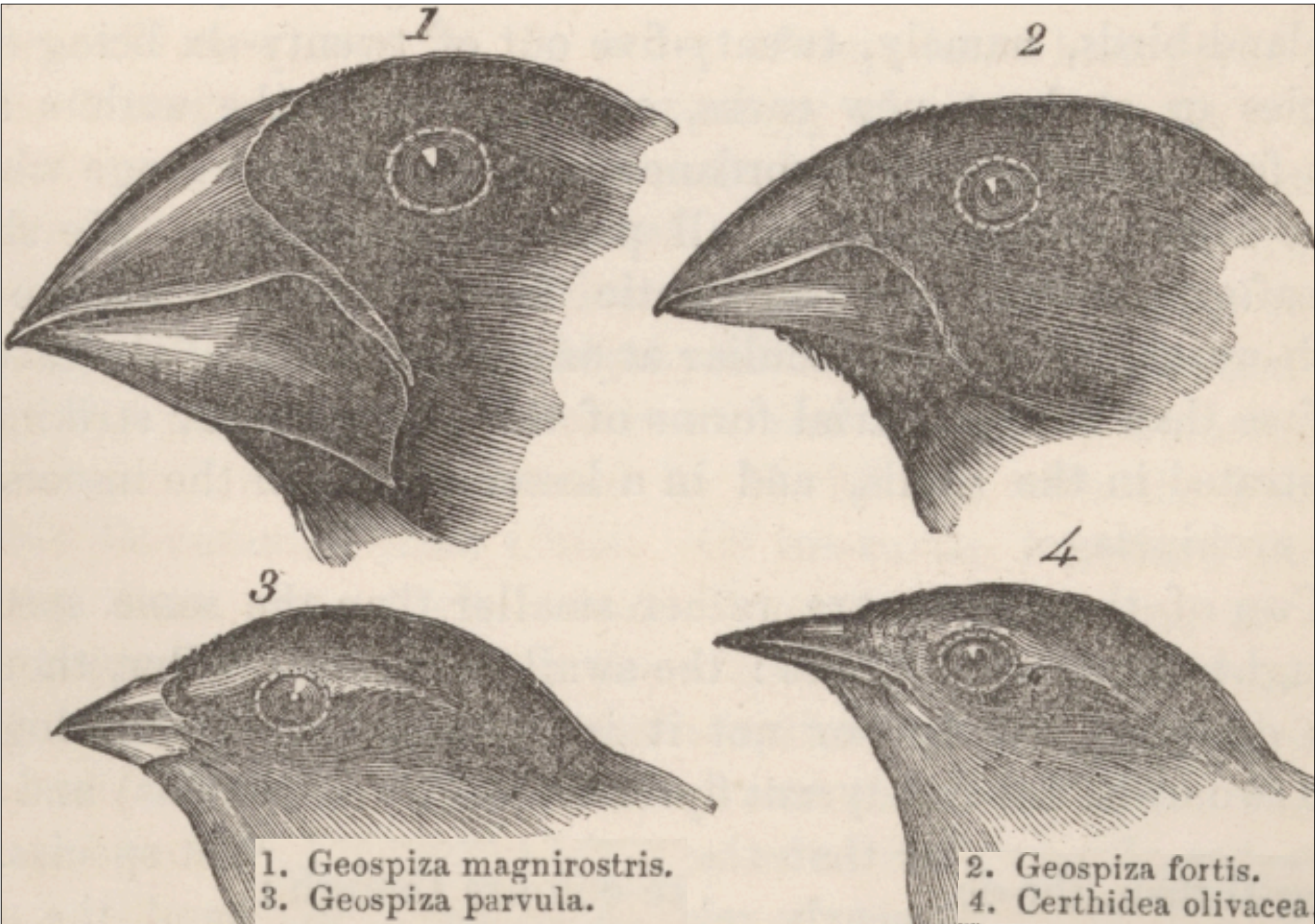
<http://www.flickr.com/photos/ole/1040941928/>



A key realization

- Knowledge on the boundaries of variation/species
- The facts
 - Large number of islands
 - Islands are spatially separated
 - Wide diversities of species (different sources of food)
 - Very close species (beyond variation)
- Start accumulating descriptions of specimens





1. *Geospiza magnirostris*.
3. *Geospiza parvula*.

2. *Geospiza fortis*.
4. *Certhidea olivacea*.

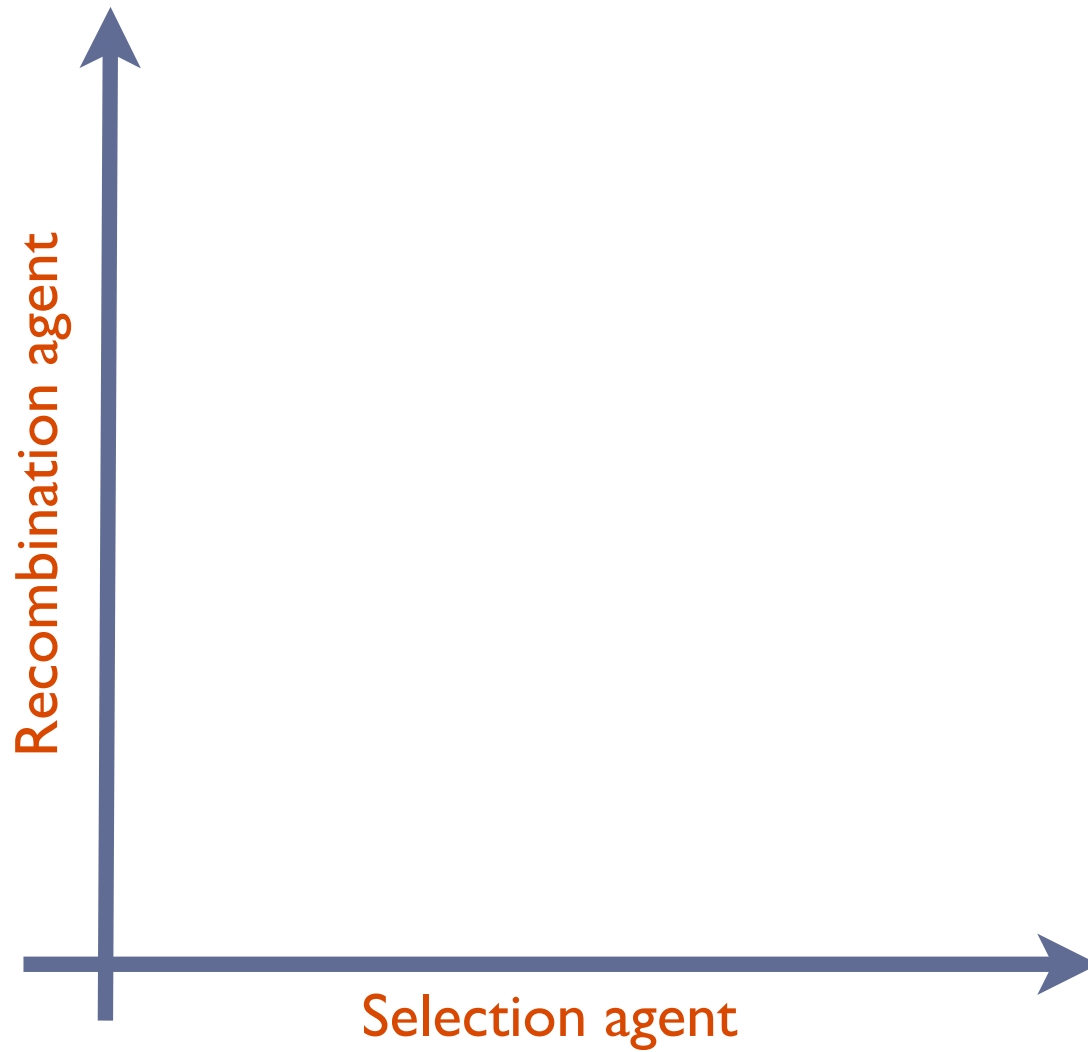


The basic building blocks

- Large variation of features
- Highly correlated to food sources
- The basic sketch of a crazy thought:
 - Assume you start with a diverse population of A s
 - Split A s and place them in different environments
 - Individual a_i features may or may not help live in the environment
 - If individual a_i features help in the environment a_i as a result has increased chances to reproduce
- The underlying distilled idea of natural selection
 - Variation will arise in a population
 - Selection by environment adaptation
- The uncanny idea: Variation can be random



A simple abstraction

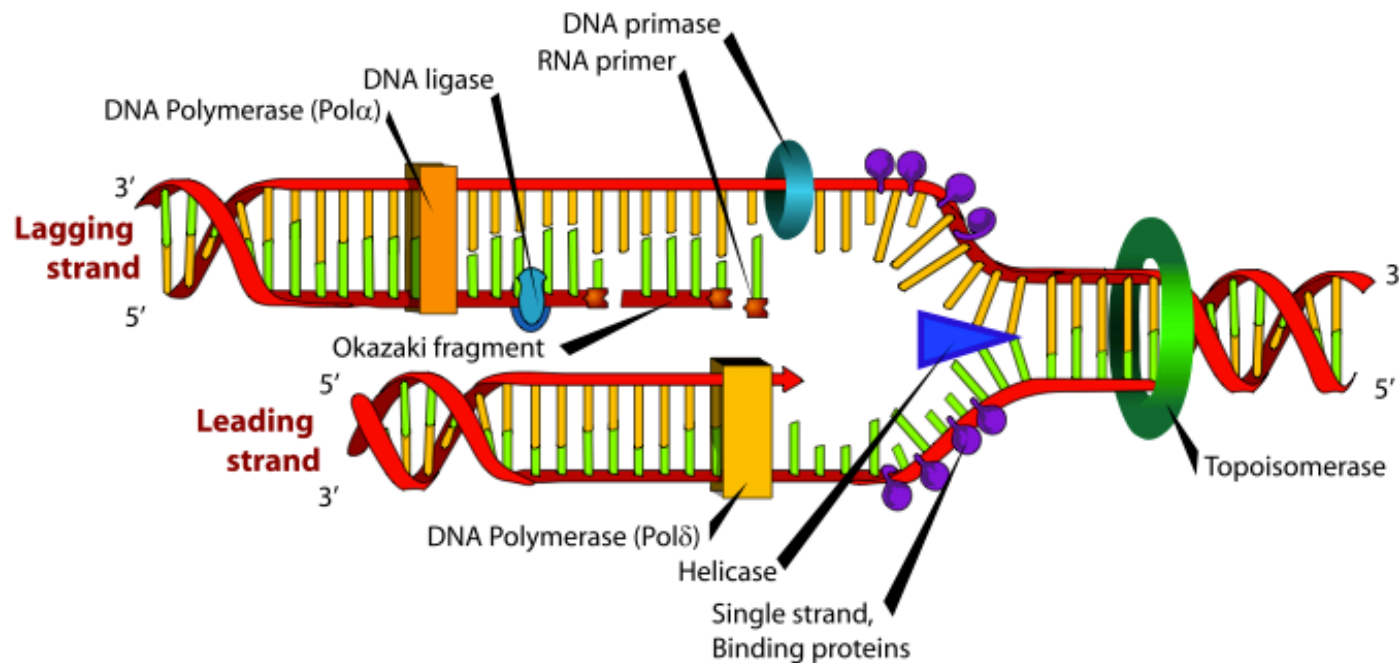
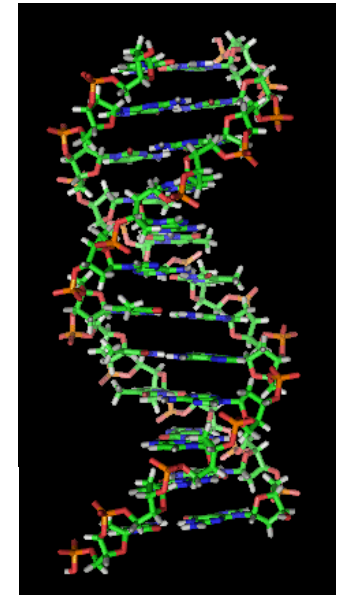


[Fast forward] ... and genomics arise



DNA and basics of reproduction

- Watson & Crick (1953)
- Four bases:
Adenine (A), cytosine (C), guanine (G), thymine (T)
- Groups in to base pairs: A-T, C-G
- Reproduction of DNA (*crossover* and *mutation*)



[Fast forward] ... and computers happen



A quick stop in the late 60s

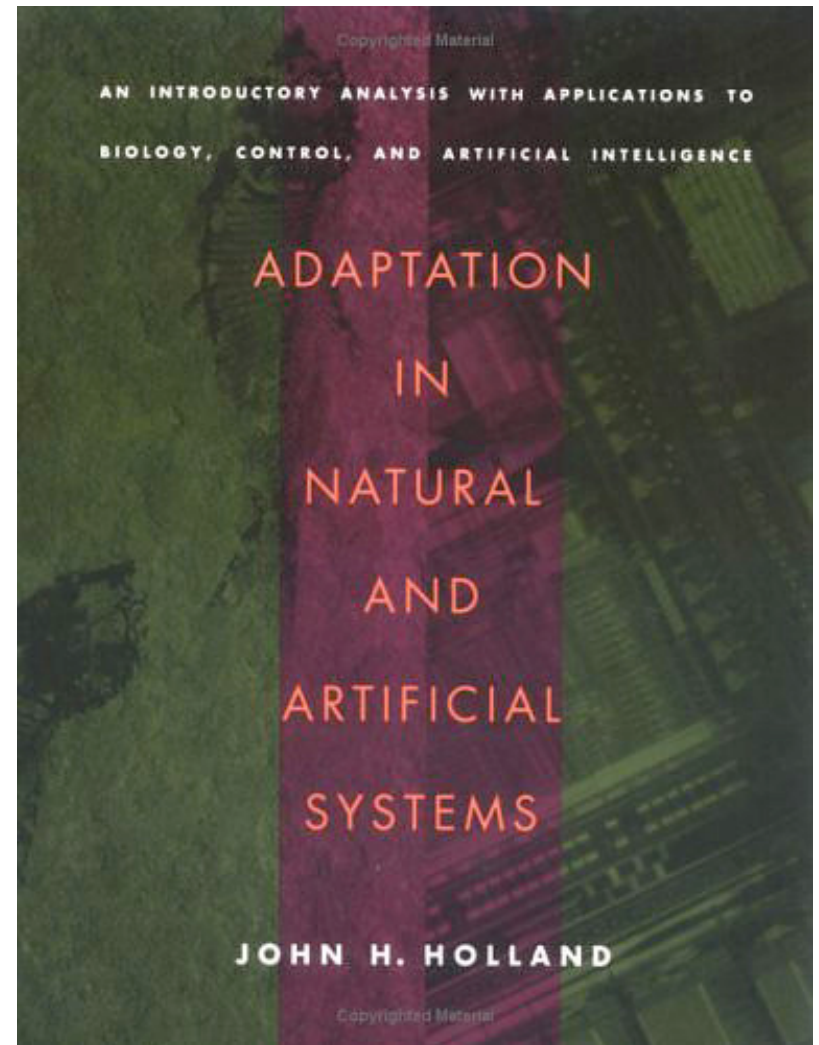
- Computers were the next cool thing
- Genetics were on the rise
- A professor at University of Michigan
- Natural evolution
 - Diverse population of individual
 - Individual a_i features may or may not help live in the environment
 - If individual a_i features help in the environment a_i as a result has increased chances to reproduce
- Artificial evolution?
 - Model of such evolutionary process
 - Use it to solve some real problem



Prof. John H. Holland

What is a genetic algorithm?

- ANAS published in 1975
- Chromosomes encode a solution to a problem
- Evaluation of fitness based on the quality solving the problem
- The process
 - Survival of the fitness
 - Mutation
 - Crossover
 - Repeat until good enough



A simple example

- Imagine you have a box with an array of switches
- 0 off, 1 on
- Depending on the switch arrangement you get some payoff
- Payoff is equivalent to the number of switches that are on
- The goal: Maximize the payoff



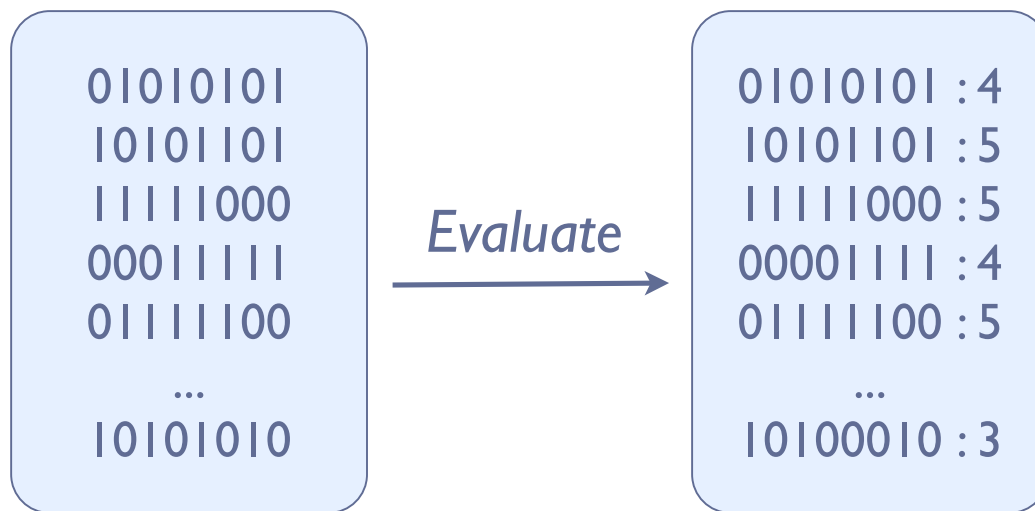
Encoding the problem

- Solutions are a string of 0s and 1s.
- We need a population

```
01010101
10101101
11111000
00011111
01111100
...
10101010
```

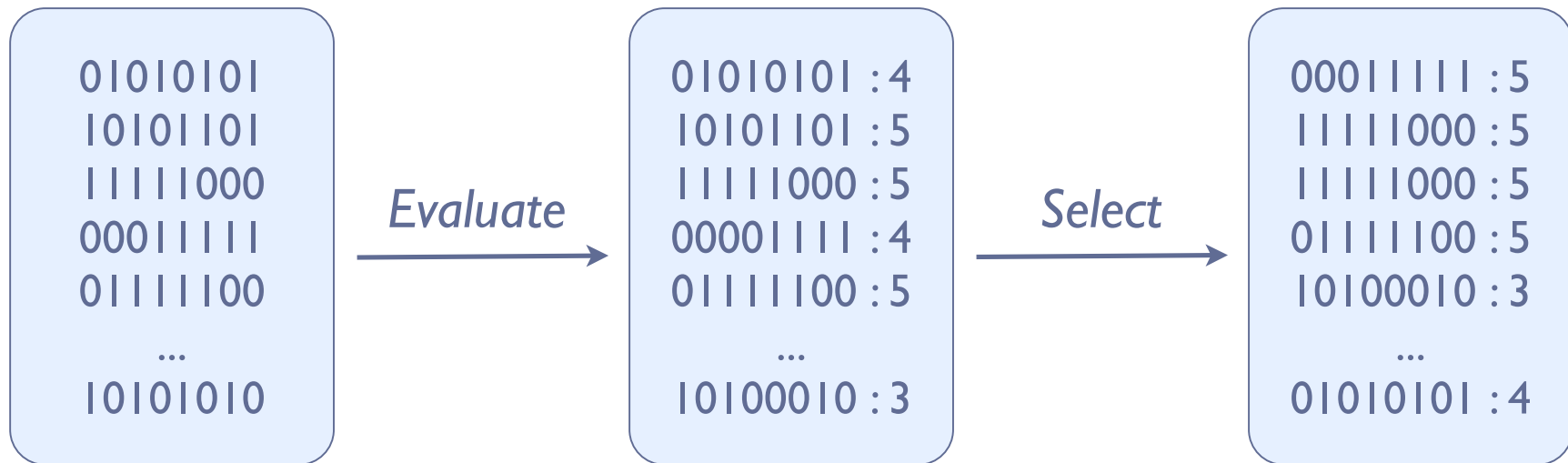
Evaluate solutions performance

- Solutions represent a solution to a particular environment
- Solutions get measured based on their performance



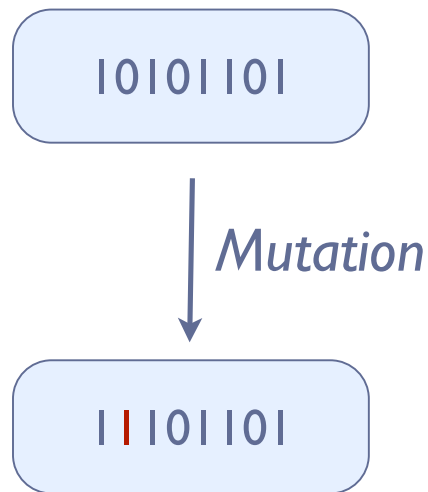
Selection based on performance

- Proportional to their performance
- Proportionate roulette wheel selection

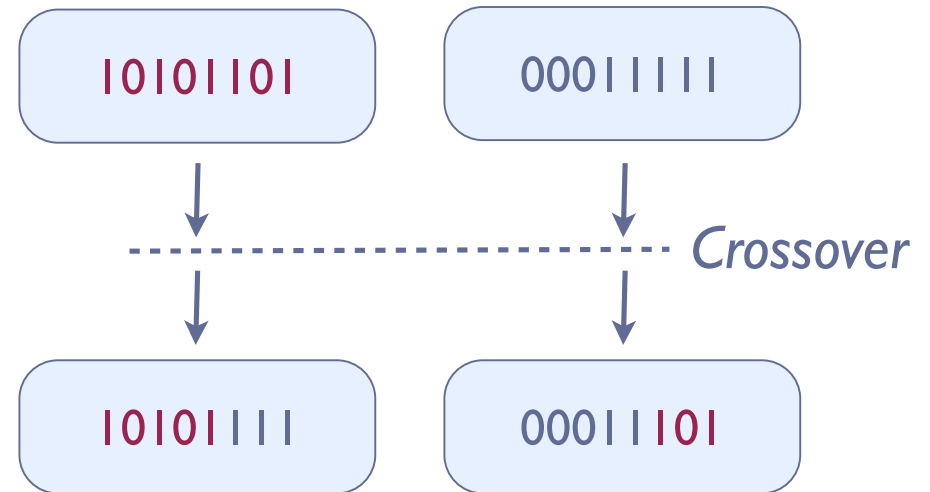


Reproduction

- Mutation introduce random changes in the solution



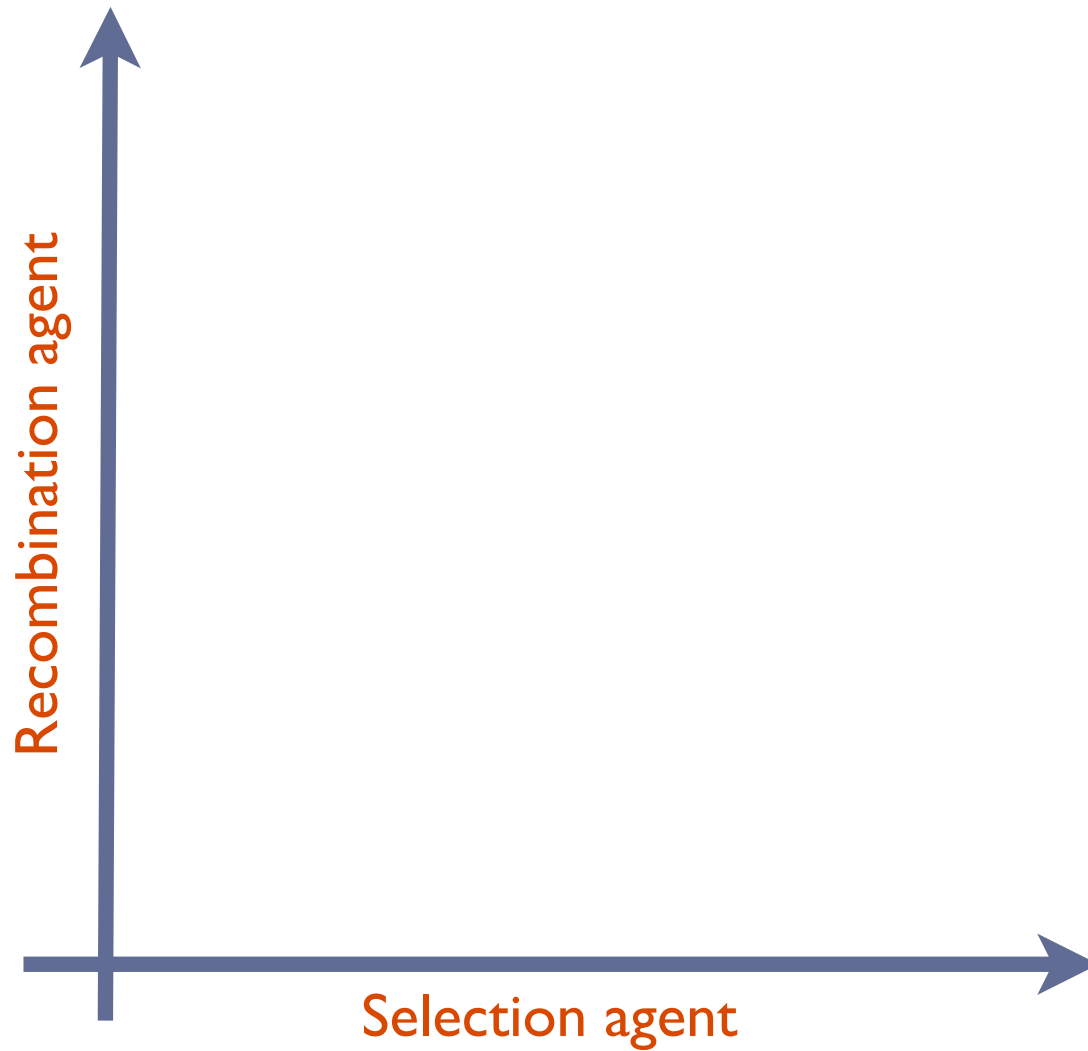
- Crossover recombines parent solutions



A nagging question

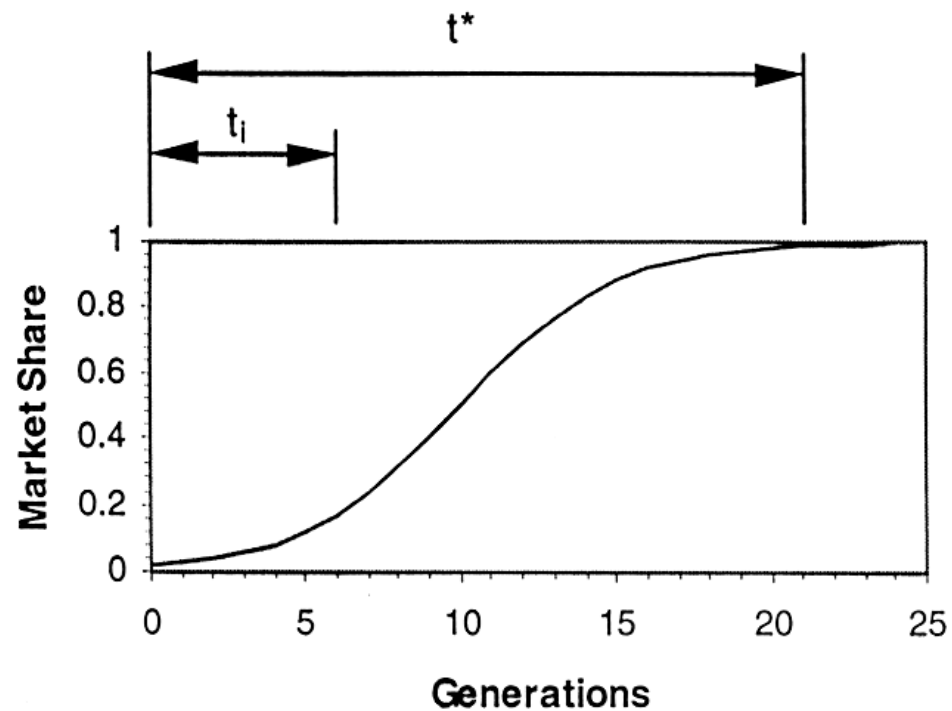
- In early 80s by David E. Goldberg
- How do uninteresting operators yield interesting behavior?
- 1983 proposed the innovation intuition:
“Genetic algorithm power like that of human innovation”
- The underlying elements:
 - Selection + mutation as hillclimbing or *kaizen*
 - Selection + crossover = *innovation*
- In another words
 - Introducing small changes and keeping them if they improve performance
 - [...] Takes two to invent anything. One makes up combinations; the other one chooses [...] (*Paul Valéry*)

Remember the simple abstraction



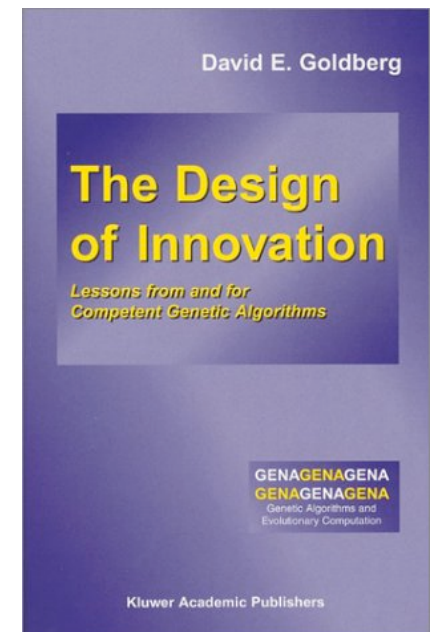
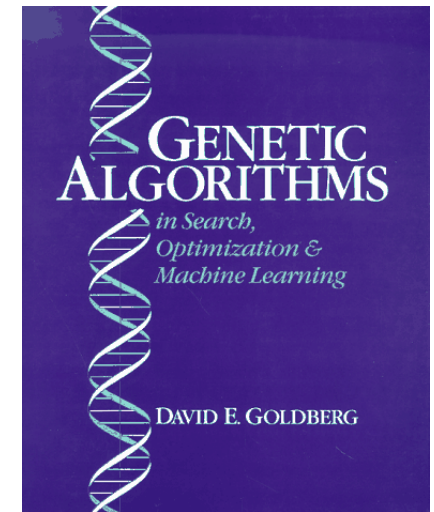
Goldberg's intuition

- Interesting behavior (emergence) is a balance
- Two components
 - Takeover time (t^*)
 - Time to innovate (t_i)



Goldberg's intuition

- A design theory for scalable and efficient GAs
 - Understanding building blocks (concepts or ideas)
 - Ensure building block supply
 - Ensure building block growth
 - Control building block speed
 - Ensure good building block decisions
 - Ensure good building block mixing
 - Know building block challengers



An interesting social insight

- GAs used for optimization, search, machine learning, ...
- But Valéry's quote also implied a social aspect to innovation
 - Solutions may be created in the vacuum, randomly, or carefully crafted
 - There is a social element of acceptance/rejection
- Paul Valéry (1871-1945) lived after Darwin published his ideas



[Fast forward] ... The 2.0 information age



Adding social to genetic algorithms

Recombination agent	computational	Standard Genetic Algorithms	Interactive Genetic Algorithms
	human	Computer Aided Design (CAD)	Human-Based Genetic Algorithms
		computational	human

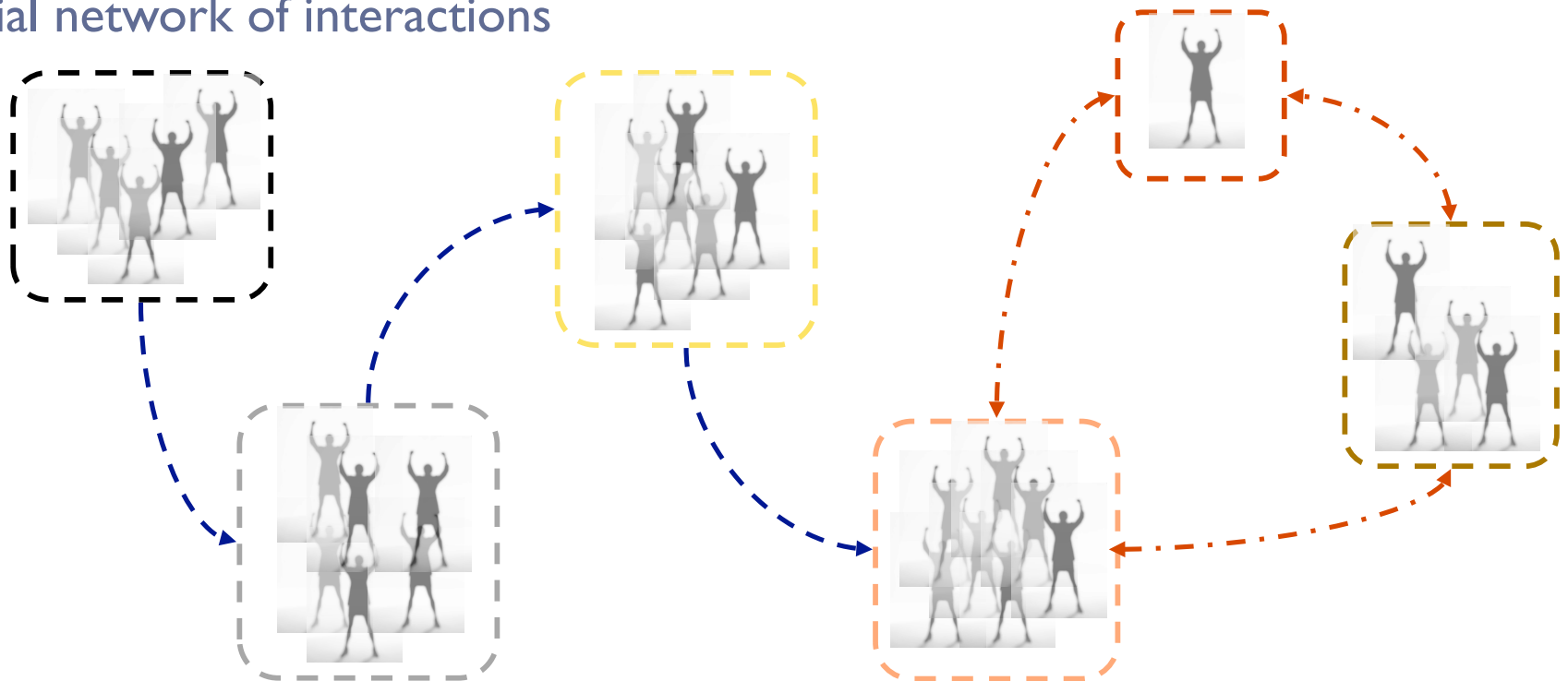
Selection agent

(Kosorukoff & Goldberg, 2002)



Social aspects of genetic algorithms

- Population-based method
- Parallel genetic algorithms
 - Groups of working units
 - Communication between groups for a common goal
 - Models for collaborative work
- Social network of interactions



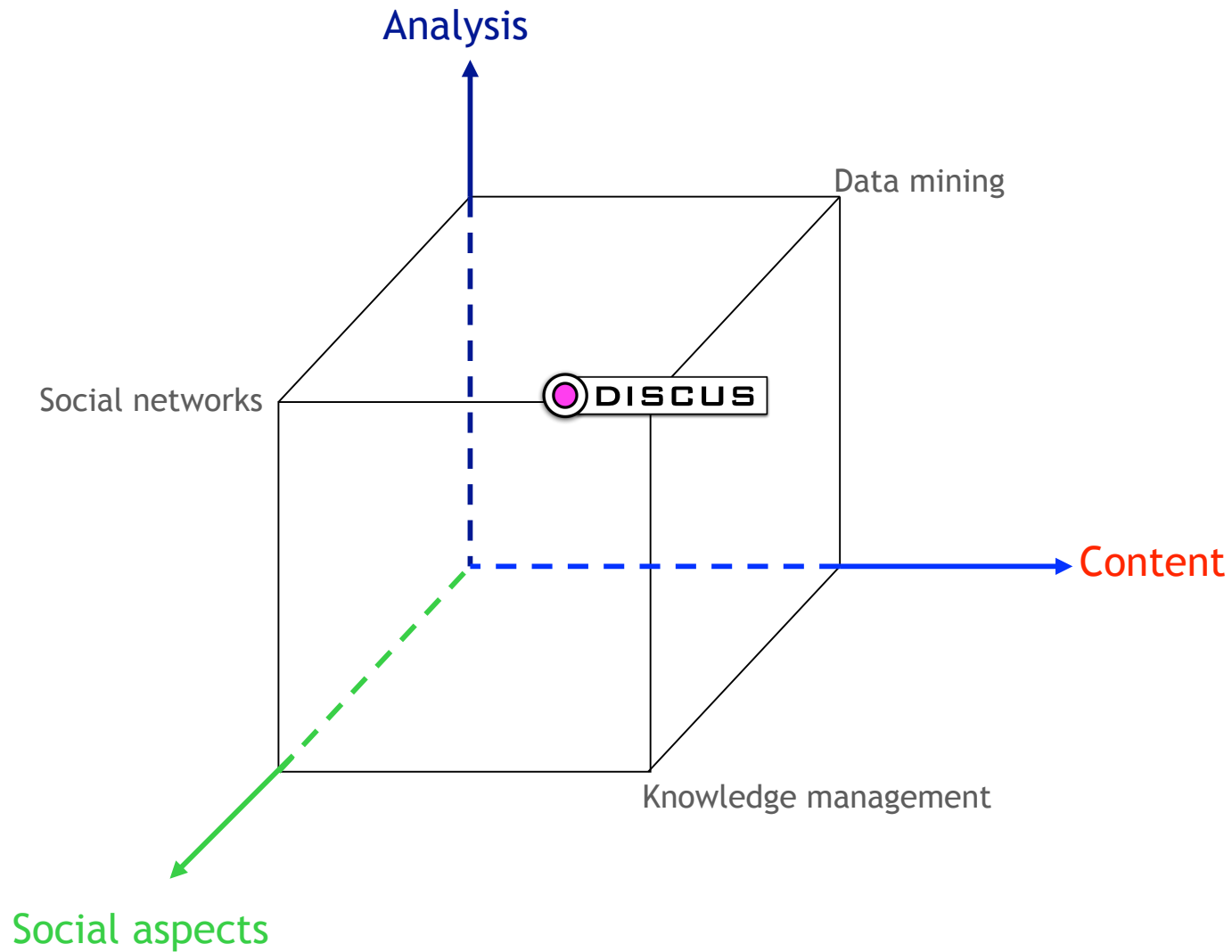
The DISCUS project

- IlliGAL & NCSA collaboration started in 2003
- Computers have become mediators of collaborations
 - Email, chat rooms, blogs, wikis...
 - A flood of available information
 - Different modes of communication
- Let's take advantage of such information
 - Logs of conversations
 - Archive of documents (email attachments, blogs, personal web pages...)
 - Human-computer interactions
 - Social aspect of the communication and collaboration

The question

- Could we help build a system using genetic algorithms/natural selection ideas that could foster innovation and creativity?
- The era of available Web 2.0 tools
 - Web portals
 - Collaboration and communication tools (message boards, chats, blogs, Twitter...)
 - Genetic algorithm (Interactive, Human-based, and Parallel GAs)
 - Chance discovery (KeyGraphs and Influence Diffusion Models)
 - Data and text mining
 - Information retrieval
 - Web services
 - Data-intensive computing

In a picture



(Llorà et al., 2004)



A simple exercise

- Imagine you are chatting with your friends
- Your ideas are bundled in messages
 - You pass them your your friends
 - They read them, think about them
 - Your friends reply to you
- Could we model it using natural selection?
 - Valéry suggest so
 - Goldberg proposed a simple model using the human-based GA analogy
- If so, Goldberg's methodology should apply



People interacting online

DISCUS Communities

DISCUS 3.0: Human-centered enterprise collaboration and innovation Welcome, admin.

HOME ANALYSIS TOOL

Next »

CELLPHONE DISCUSSION

27 April 2007. Posted by saru.

Posted in Cellphone | RSS feed **40** Comments »

Comment by user1 on 2007-04-27 06:07:27 | [Edit this](#) | [Reply to this comment](#)

Let us start discussion. What is the benefit of cell phones in your daily life? And, what is the dissatisfaction of it?

Comment by user2 on 2007-04-27 06:07:59 | [Edit this](#) | [Reply to this comment](#)

benefit: makes my life easier. everywhere, anywhere, i can take important calls that i wouldn't want to miss. also there are some interesting features that cellphone has, like calendar, alarm and everything. since i always bring cellphone everywhere, it is nice to be able to do all my scheduling and stuff there too.dissatisfaction: maybe the fact that sometimes it is not as reliable as i would like it to be. and there is not enough memory.

Comment by user3 on 2007-04-27 06:08:24 | [Edit this](#) | [Reply to this comment](#)

I wish the calendar feature on phones was more integrated with a computer use so that I could create a calendar on my computer and then transfer it over to my phone.

Comment by user4 on 2007-04-27 06:08:45 | [Edit this](#) | [Reply to this comment](#)

hmm... good point. i think it is important for cellphone to be able to interact with other electronical devices, such as computer, digital camera, electronic organizer, etc.

Comment by user1 on 2007-04-27 06:09:31 | [Edit this](#) | [Reply to this comment](#)

Going back to my dissatisfaction in my cell phone: I wish it were cheaper to use the internet services mostly so that I could check for important emails when I don't have access to a computer when I need one.

DISCUSSIONS

- » Cellphone
- » discuss_dev
- » Normal
- » ozawa_watabe
- » watabe
- » yuichi_morito

RECENT ENTRIES

- » DISCUS3の今後の方向性に関して
Poll Results:
» Where is the worst?
Poll Results:
» Where is the best?
» 東京のお勤めの観
» 光名所は？ [1人]
- » 東京のお勤めの観
» 光名所は？ [3人]

POLLS

- » Polls

META

- » Site Admin
- » Logout



Concepts as building blocks

DISCUS Analytic Tools

RANKING : TERM	TOP RANKED SENTENCES
1 : Twitter	<i>After Twitter rejected the company's acquisition offer, it was no coincidence that real time updates, friend following and improved activity streams were incorporated into Facebook's redesign.</i>
2 : Facebook	
3 : companys	
4 : billion	<i>This new round of funding suggests that Twitter is taking a page out of the Facebook playbook and attempting to increase its abilities as a platform.</i>
5 : attempting	
6 : platform	<i>If Twitter is attempting to become the "pulse of the planet" as suggested by the widely criticized leak of the company's internal documents, then perhaps \$1 billion dollars isn't entirely off.</i>
7 : valuation	
8 : million	
9 : announcing	
10 : streaming	<i>And like Twitter, while we've seen huge traffic numbers, Facebook only yesterday announced becoming cash-flow positive.</i>
11 : funding	
12 : round	<i>In late May, Facebook had a valuation of \$10 billion dollars based on a \$200 million investment from Digital Sky Technologies.</i>
13 : real	
14 : ability	
15 : investment	
16 : rejected	
17 : acquisition	
18 : coincidence	
19 : updates	
20 : friend	
21 : incorporated	
22 : redesign	
23 : taking	
24 : page	
25 : playbook	
26 : pulse	
27 : planet	
28 : widely	
29 : criticized	

Done zotero

DISCUS Analytic Tools

Graph Controls

DISCUS

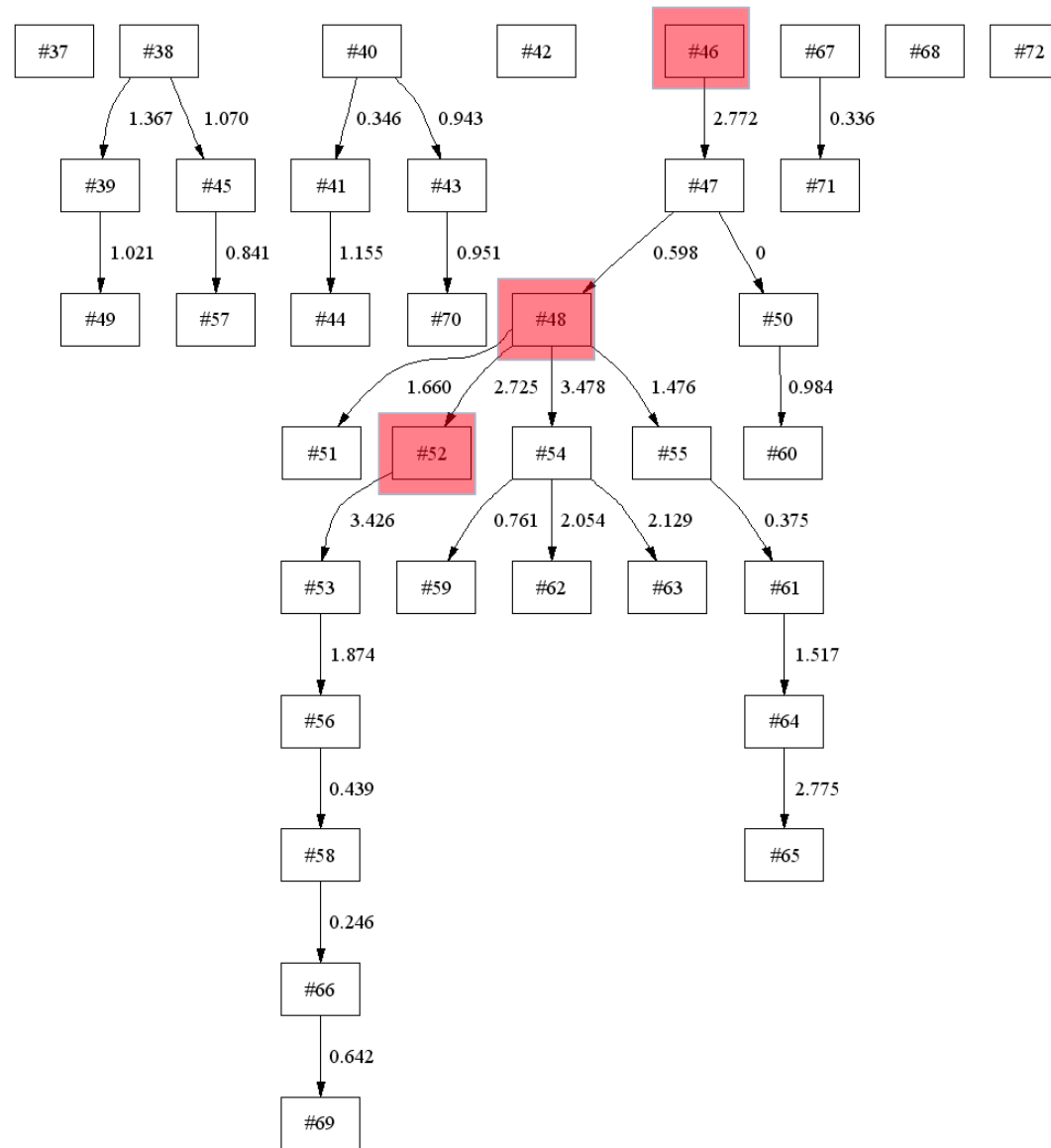
Copyright by UIUC-ILLIGAL-NCSA 2003-2006.

Applet org.idiscus.tools.visualization.prefuse.KeyGraphApplet started zotero

(Llorà et al., 2004)



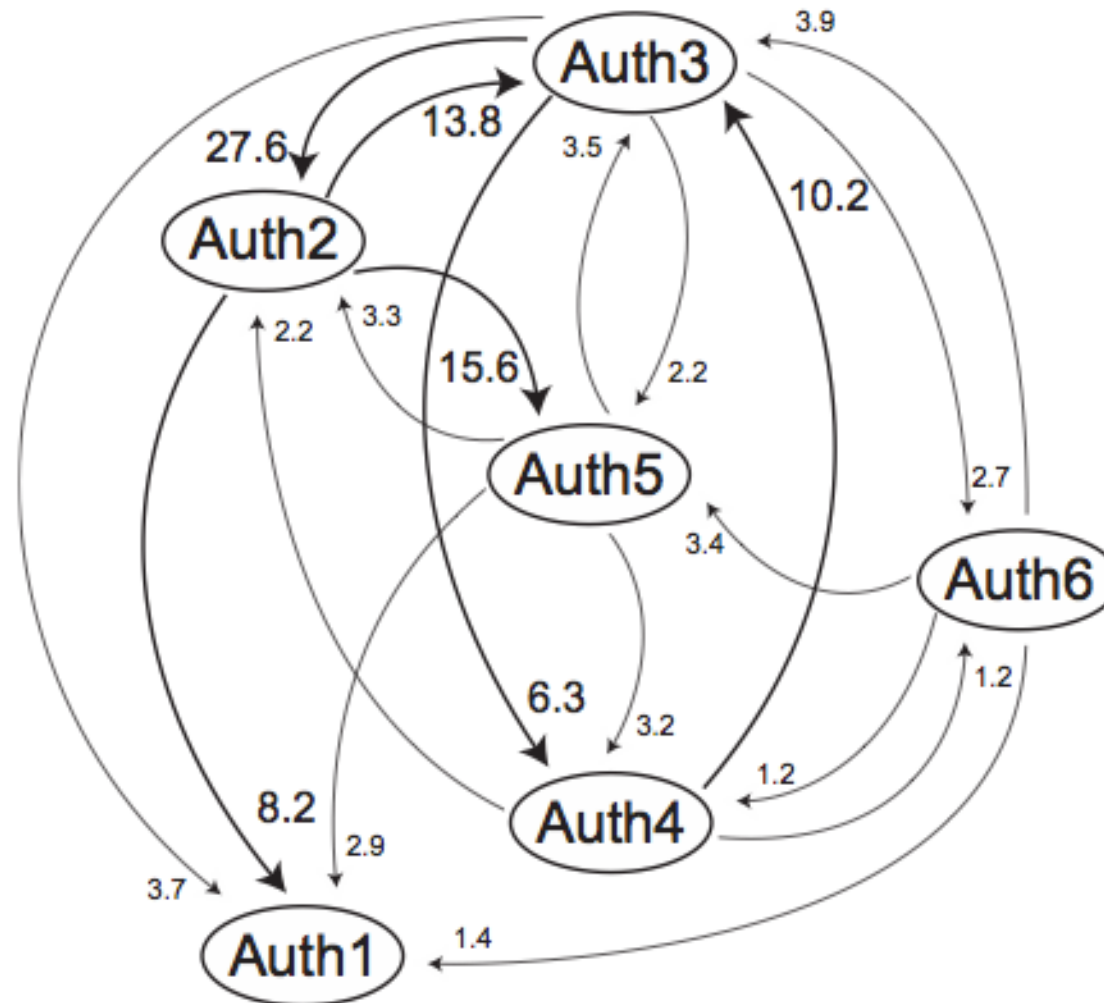
Things you can measure: Concept diffusion



(Llorà et al., 2005)



Social acceptance and roles



(Yasui et al., 2006)

Everything together

0%

discussion

Please generate make ideas about "future cell phone". The ideas are what kind of service, gadget, social phenomenon will emerge.
ex) Apple will release thin, multitouch cellphone.

Brainstorming rules

- Focus on quality
- No criticism
- Unusual ideas are welcome
- Combine and improve ideas

Posts

Comments

Discussion




comment by saru on 2008/01/20 12:21 | [reply to this](#) |

Going back to my dissatisfaction in my cell phone: I wish it were cheaper to use the internet services mostly so that I could check for important emails when I don't have access to a computer when I need one.

comment by saru on 2008/01/20 12:21 | [reply to this](#) |

The benefits of cell phones in my daily life are that: I can always be reached, I never really feel out of the loop, I do not have to delay asking someone a question til I see them or am near a computer, and that I just feel more comfortable with it then without. The only dissatisfaction I might get from it is having people that I would rather not talk to call me but this can be easily avoided by just not answering or sending to voicemail.

Users

-  tamdai
online
-  kazu
offline
-  saru
online

Phase

- » [discussion](#)

META

- » [Site Admin](#)
- » [Logout](#)

Navigation bar

(Saruwatari et al., 2008)



Everything together

Progress bar 4%

Initial Idea Generation

Please generate 5 ideas about "future cell phone". The ideas are what kind of service, gadget, social phenomenon will emerge.
ex) apple will release thin, multitouch cellphone.

Input form

Initial Ideas

Posted on 2008/01/19 13:38

Be implanted to human body

Generated ideas

Progress bar 5%

Idea Crossover

Please generate make new ideas from following two ideas.

Association

Be implanted to human body	+	Personal authentication tool
Biometrics		
Be implanted to human body	+	communication speed is as same as current WiFi
Input form		
Be implanted to human body	+	Have interpreter

Stimulus

Stimulus

Progress bar 4%

Idea Selection

Please vote 5 ideas which you prefer.

Remaining number of vote is 4

vote button

Be implanted to human body

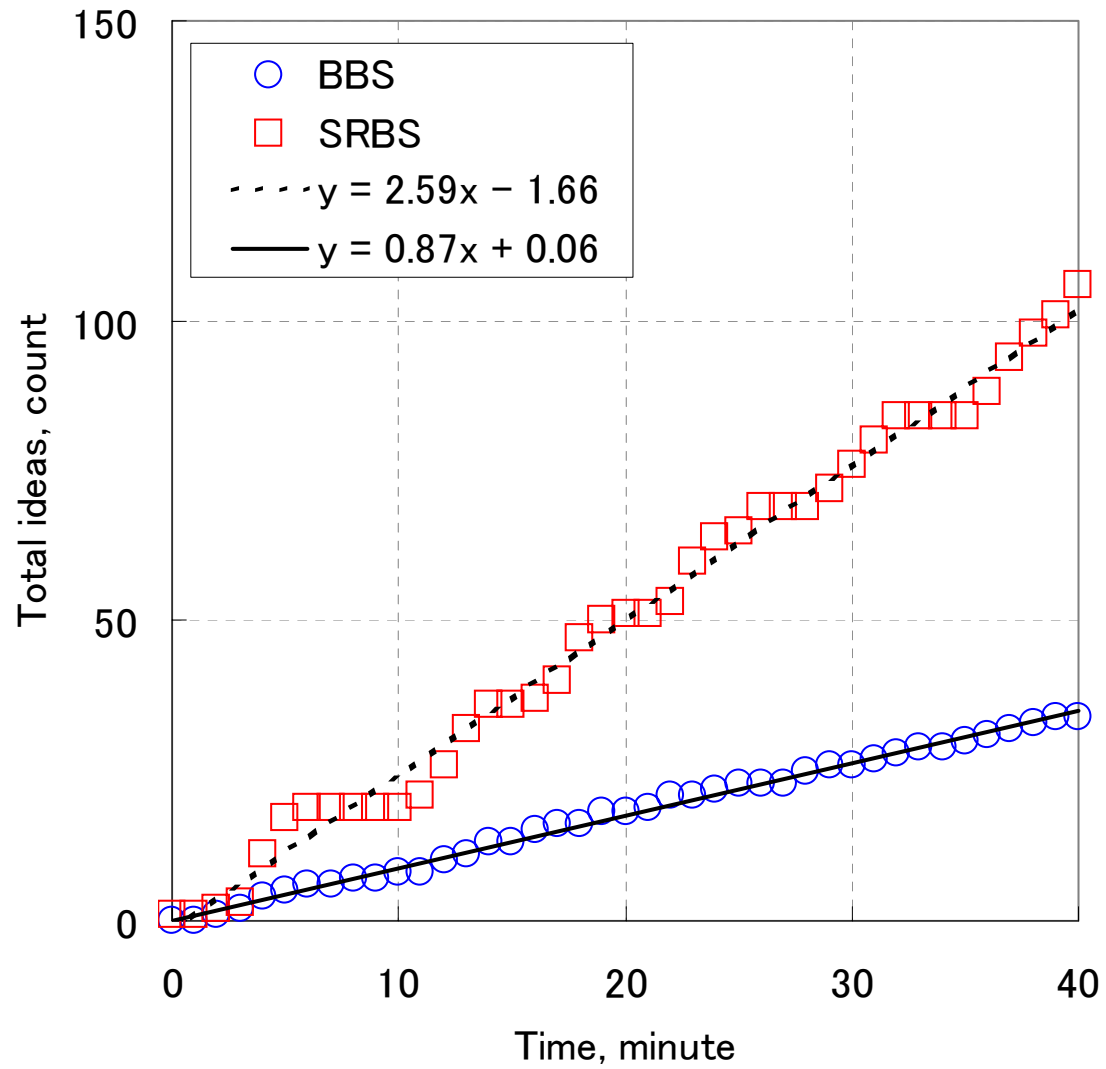
communication speed is as same as current WiFi

Have interpreter

(Saruwatari et al., 2008)



In action



(Saruwatari et al., 2008)



Other possible usages

- Tracking influence on social media (e.g. Twitter)
- User preference modeling
- Topic overlap
- Topic/trend dynamics (how conversation is changing on the fly)
- Blogosphere and company message misalignment
- ...

... and beyond



SW development and Darwin's ideas?

- Data deluge, processing large volumes of data starting to become the norm, not the exception
- Another nagging question back on 2007
- Could we use the basic design principles by Goldberg to build software?
- A subset of those could have the answer
 - Understanding building blocks (concepts or ideas)
 - Ensure building block supply
 - Ensure building block growth
 - Ensure good building block mixing
- Also
 - Exploiting parallelism efficiently is becoming a more pressing issue
 - SW is a social activity



A little thinking

- Understanding building blocks (concepts or ideas)
 - Basic units of data processing
 - Building blocks should be easy to group and swap
- Ensure building block supply
 - Create a basic set of building block
- Ensure building block growth
 - Provide a set of tools for developers to add more building blocks
- Ensure good building block mixing
 - Facilitate the reproduction of solutions (programs)
 - Provide means to make those recombination possible



Modeling Meandre

- The basic outline
 - Data-flow execution paradigm
 - Semantic-web driven
 - Web oriented
 - Facilitate distributed computing
 - Support publishing services
 - Promote reuse, sharing, and collaboration
- Fall 2007 Meandre was adopted as the data-intensive platform for a collection of projects
- More information at <http://seasr.org/meandre>



Data flow execution in Meandre

- A simple example $c \leftarrow a+b$
- A traditional control-driven language

a = 1

b = 2

c = a+b

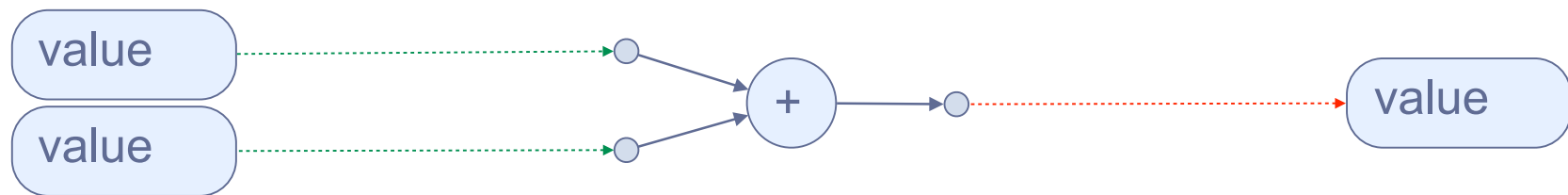
- Execution following the sequence of instructions
- One step at a time
 - $a+b+c+d$ requires 3 steps
 - Could be easily parallelized



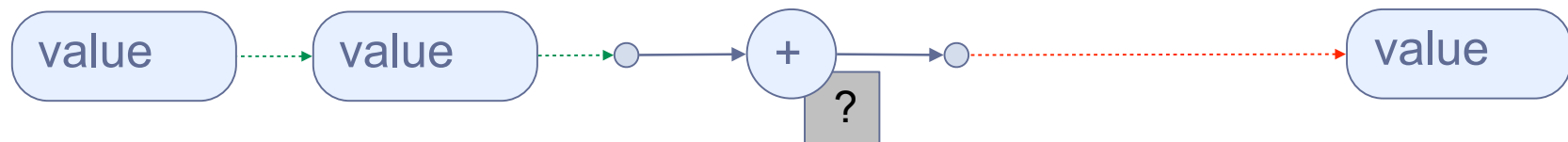
Data flow execution in Meandre

- Data flow execution is driven by data
- The previous example may have 2 possible data flow versions

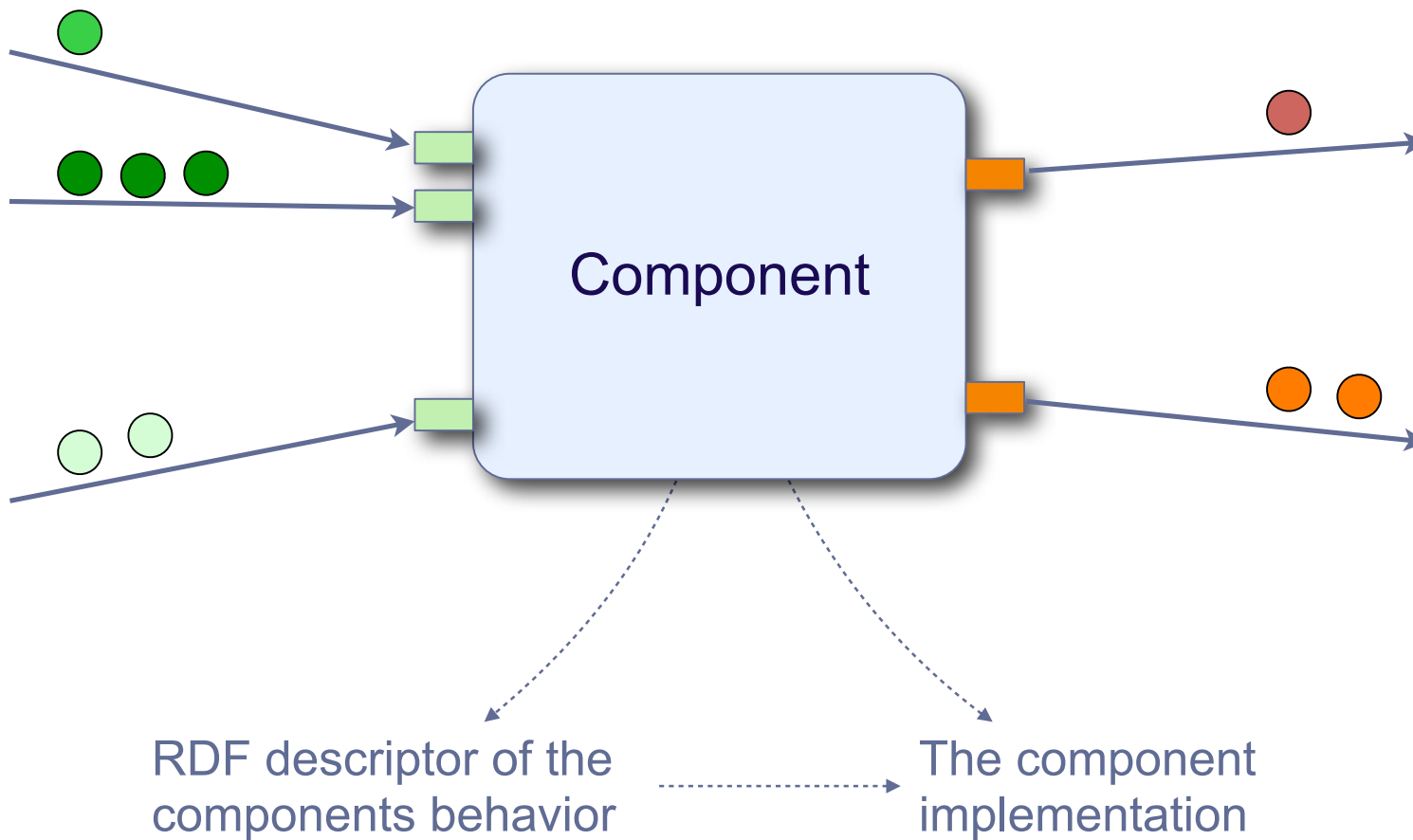
Stateless data flow



State-based data flow

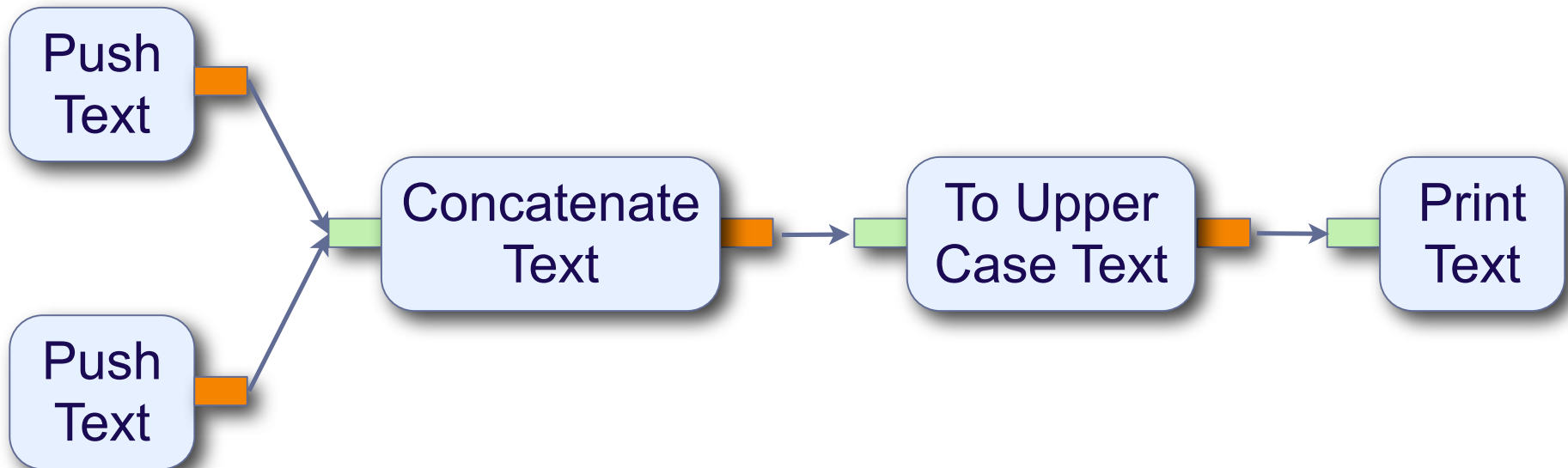


The basic building blocks: Components



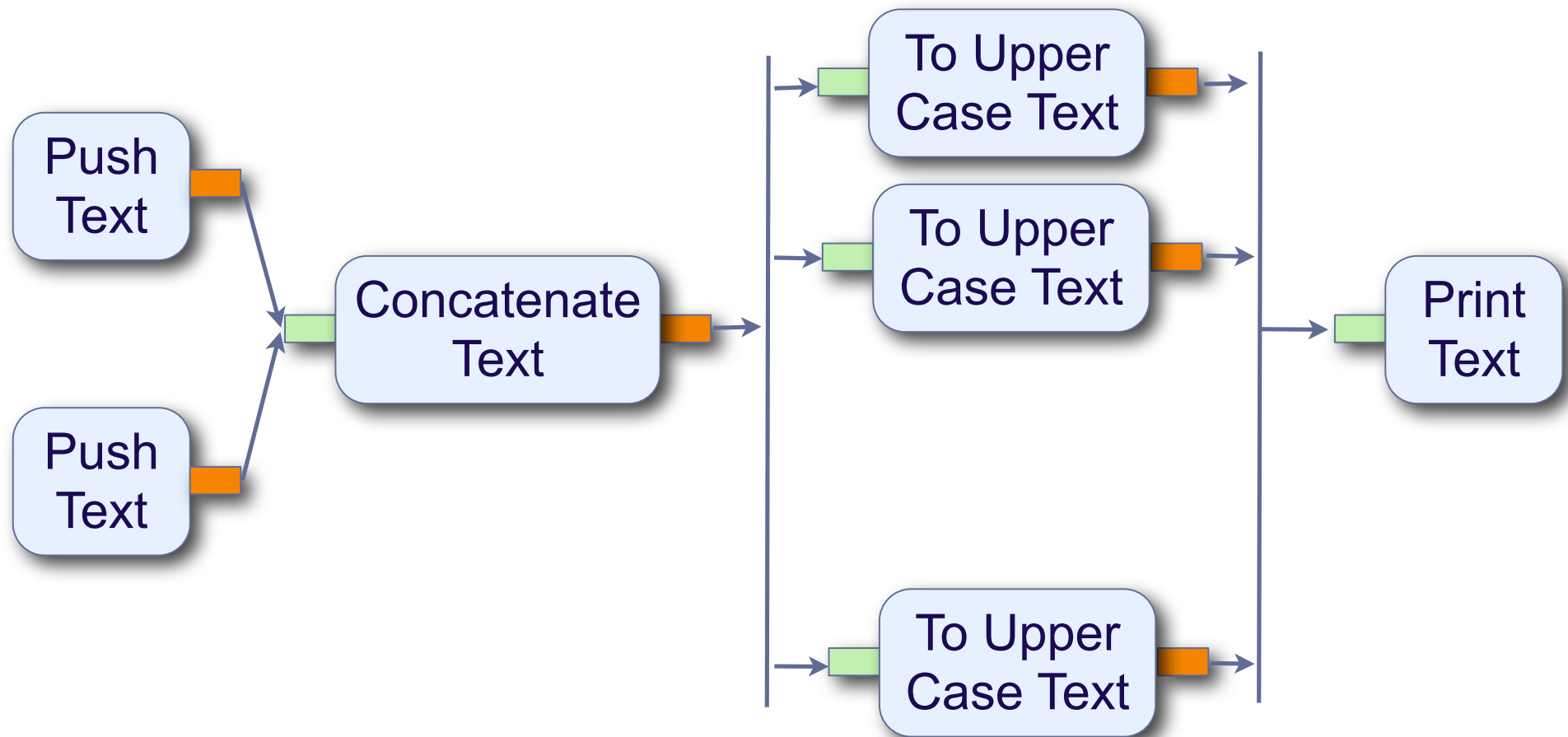
Go with the flow: Creating complex tasks

- Directed multigraph of components creates a flow



Automatic parallelization: Speed and robustness

- Meandre ZigZag language allow automatic parallelization



Visual building block recombination

The screenshot displays the Meandre-Workbench interface, a visual programming environment for text processing. The main workspace shows a workflow with the following components:

- Input Text
- Universal Text Extractor
- To Lowercase
- OpenNLP Sentence Detector
- OpenNLP Sentence Tokenizer
- Push Text
- Universal Text Extractor
- Concatenate Text
- OpenNLP Tokenizer
- Token Filter
- HITS summarizer
- Sentences to text
- Concatenate Text
- HTML Fragment Maker
- HTML Viewer
- Tokens To Text

The workflow starts with 'Input Text' and 'Push Text' feeding into 'Universal Text Extractor'. The output goes through 'To Lowercase', 'OpenNLP Sentence Detector', and 'OpenNLP Sentence Tokenizer'. Another 'Push Text' feeds into another 'Universal Text Extractor', which then feeds into 'Concatenate Text'. The output of 'Concatenate Text' goes to 'OpenNLP Tokenizer', which then feeds into 'Token Filter'. The output of 'Token Filter' goes to 'HITS summarizer', which then feeds into 'Sentences to text'. The output of 'Sentences to text' goes to 'Concatenate Text', which then feeds into 'HTML Fragment Maker', and finally 'HTML Viewer'. The 'Tokens To Text' component is highlighted in the workflow.

The interface includes a 'Repository' on the left with a list of components and flows. The 'Details' panel on the right shows the properties of the selected component:

Name	Value
count	-1
debug_level	info
header	true
ignore_errors	false
message	Top Tokens
offset	0

The 'Description' panel on the right provides details for the 'Tokens To Text' component:

Tokens To Text
<http://seasr.org/flows/custom-hits-summarizer/instance/tokens-to-text/2>

Tags: tokenizer, text, tools, counting, ser

Creator: Xavier Liora

Date: Mon Jun 08 18:06:13 GMT-500 20X

Class: org.seasr.meandre.components.an

Description:
Given a collection of tokens, this compone converts it into text.

Rights:
UofNCSA

Inputs:
tokens

Where is natural selection?

- More DISCUS-like evolutionary modeling
- Variations (components) in a population of solutions (flows)
- Selection by usage/adoption
- Recombination of flow available (encapsulation)



The take away message



Basic food for thought

- The underlying distilled idea of natural selection
 - Variation will arise in a population
 - Selection by environment adaptation
- Powerful modeling tool
 - Problem solving (optimization and search)
 - Machine learning
 - Human social interaction modeling
 - Defining new ways to explore computer programming
 - ...
- Don't be afraid to try and explore



More information

- Illinois Genetic Algorithms Laboratory
 - <http://www.illigal.uiuc.edu/web>
- The DISCUS project
 - <http://www.i-discus.org>
- The Meandre project
 - <http://seasr.org/meandre>
 - <http://seasr.org>
- National Center for Supercomputing Applications
 - <http://www.ncsa.illinois.edu>
- Homepage
 - <http://www.xavierllora.net>
- Some related Twitter users
 - People: @xllora, @deg511
 - Labs and projects: @illigal, @projectmeandre, @seasrproject



From Galapagos to Twitter: Darwin, Natural Selection, and Web 2.0

Xavier Llorà

National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
Urbana, Illinois, 61801

xllora@ncsa.illinois.edu

<http://www.ncsa.illinois.edu/~xllora>