Building Business Models with Big Data

Prof. Stephen K Markham, Ph.D.
Department of Management, Innovation & Entrepreneurship
Poole College of Management
North Carolina State University
The Valley of Death

Level of Development

Resources

Idea Generation Resources

Commercialization Resources
Complete the sequence:

• Fred, . . .
Complete the sequence:

• Fred, Daphne, . . .
Complete the sequence:

- Fred, Daphne, Velma, . . .
Complete the sequence:

• Fred, Daphne, Velma, Shaggy, . . .
Complete the sequence:

- Fred, Daphne, Velma, Shaggy, Scooby Doo
Complete the sequence:

- **Fred, Daphne, Velma, Shaggy, Scooby Doo**
  - Famous for solving mysteries “Scooby-Scooby-Doo”
  - Called together as a team to solve a mystery “Reah”
  - They all wanted to go in different directions “Ruh?”
  - Would not work together “Rhut Rowh”
    - Each had different parts but not coordinated
  - The bad guy got the upper hand “Ruht-Rowh Raggy”
  - Late one night – they were all captured . . .
    - . . . except Scooby and Shaggy “Zoiks”
  - They managed to break their friends free “Reah”
    - They pulled together as a team
    - Formed a plan
    - Used everyone's unique abilities
    - Increased morale
  - Caught the bad guy “Scooby Scooby-Doo”
  - Saves the day “Scooby Laugh”
It’s the question; not the technology…

Crossing the Valley of Death

Level of Development

Resources

Idea Generation Resources

“Valley of Death”

Commercialization Resources
Business Model

A business model describes the rationale for how an organization creates, delivers and captures value

http://www.businessmodelgeneration.com/canvas
Business Model Canvas

<table>
<thead>
<tr>
<th>KEY PARTNERS</th>
<th>KEY ACTIVITIES</th>
<th>VALUE PROPOSITION</th>
<th>CUSTOMER RELATIONSHIPS</th>
<th>CUSTOMER SEGMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KEY RESOURCES</th>
<th>CHANNELS</th>
<th>COST STRUCTURE</th>
<th>REVENUE STREAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It's the question; not the technology…
## Value Proposition

<table>
<thead>
<tr>
<th>Structure</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Sentence:</strong></td>
<td></td>
</tr>
<tr>
<td>For (target customer)</td>
<td>Doctors treating patients with borderline cholesterol</td>
</tr>
<tr>
<td>Who (statement of the need or opportunity)</td>
<td>Need to know how aggressively to treat cholesterol levels</td>
</tr>
<tr>
<td>The (product or service name)</td>
<td>The LipoProfile measures the number of lipoprotein particles</td>
</tr>
<tr>
<td>That (statement of benefit)</td>
<td>Provides a more accurate assessment of coronary heart disease risk</td>
</tr>
<tr>
<td><strong>Second Sentence:</strong></td>
<td></td>
</tr>
<tr>
<td>Unlike (primary competitive offering)</td>
<td>Lipid panels that only measure the amount of cholesterol</td>
</tr>
<tr>
<td>Our offering (statement of primary differentiation)</td>
<td>Uncovers the unseen risk that an elevated number of lipoprotein particles represent</td>
</tr>
<tr>
<td><strong>Third Sentence:</strong></td>
<td></td>
</tr>
<tr>
<td>Therefore, we request</td>
<td>Routinely use the LipoProfile</td>
</tr>
</tbody>
</table>
It’s the question; not the technology…

The Long Tail

- Short Head
  - Blockbusters
  - Top 40
  - Widely popular
  - Short-lived
  - Narrow scope

- Long Tail
  - Blockbusters in a niche
  - Narrowly popular
  - Popular in the past
  - Good, but not great content
  - D-list content

Popularity of Individual Titles

Content Titles

Narrow - Infinite
It's the question; not the technology…

3 Types of BUSINESS:

- Bundled
- Unbundling
- Unbundled!

Best to focus on 1!
Multisided Business Model

Value Chain of different industries intersect. E.g.,
Delivery of ad-funded content:
1. Content provider see to deliver discounted content to end users
2. Advertisers fund the discount
3. Advertisers, Content Providers and End Users Pay Telco for providing support for the transaction
Free Business Model
It’s the question; not the technology…

- **Closed Business Model**
  - Internal Development Costs
  - Market Revenue

- **Open Business Model**
  - Internal and External Development Costs
  - Market Revenue

- **New Revenues**
  - Sale/Divest
  - Spinoff
  - License

Cost and time savings from leveraging external development
Building Business Models with Big Data
The “Challenge” of Big Data

- What is It?
- Management experienced in intuitive decision making
- Too much information
- Unable to articulate actionable inquiry
What Happens in an Internet Minute?

639,800 GB of global IP data transferred
20 New victims of identity theft
47,000 App downloads
204 million Emails sent
$83,000 In sales
61,141 Hours of music
20 million Photo views
320+ New Twitter accounts
3,000 Photo uploads
100,000 New tweets
135 Botnet infections
1,300 New mobile users
100+ New LinkedIn accounts
277,000 Logins
6 million Facebook views
2+ million Search queries
6 New Wikipedia articles published
And Future Growth is Staggering

Today, the number of networked devices = the global population
By 2015, the number of networked devices = 2x the global population
In 2015, it would take you 5 years to view all video crossing IP networks each second
Big Data ≠ Knowledge
Big Data Background

• 90% of the world’s data was created in the last 2 years (IBM)
  – How have your decision making models changed in the last 24 months?

![Diagram showing increase in data amount and signals vs noise over time.]

Amount Of Data

Time

Noise

Signal

Amount of data that the firm can process
What’s Changed?

• Commodity priced computing
  • Amazon “cloud”
• Massive file system storage and retrieval technology
  • Google, MapReduce, and Hadoop
• Bandwidth
  • Grey cable
• Smart devices
  • Records are everywhere
# Types of Big Data

<table>
<thead>
<tr>
<th>Big Data Characteristics</th>
<th>Aggregation $(+ - \times \div)$</th>
<th>Isolation (NLP UTA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variety</td>
<td><em>Structured</em></td>
<td><em>Unstructured</em></td>
</tr>
<tr>
<td>Velocity</td>
<td>Real Time</td>
<td>Archival</td>
</tr>
<tr>
<td>Volume</td>
<td>Additive</td>
<td>Sift and Integrate</td>
</tr>
<tr>
<td>Veracity</td>
<td>Auditable</td>
<td>Triangulate</td>
</tr>
</tbody>
</table>
Structured Data Analytics
It's the question; not the technology…

Unstructured Data Analytics
It’s the question; not the technology…
Pick the “Right” Tool

Structured

• What happened?
• What did we sell?
• What else do they buy?
• What site did they visit?

Unstructured

• What does the customer want?
• Where is our market?
• What features are requested?
• Who do they trust?

Need to Understand Problem Type
Analyzing Ambiguity

Actual Newspaper Headlines

- Iraqi Head Seeks Arms
- Juvenile Court to Try Shooting Defendant
- Teacher Strikes Idle Kids
- Kids Make Nutritious Snacks
- Local HS Dropouts Cut in Half
- British Left Waffles on Falkland Islands
- Red Tape Holds Up New Bridges
- Clinton Wins on Budget, but More Lies Ahead
- Ban on Nude Dancing on Governor’s Desk
Natural Language Processing Requires a System

• Natural language is:
  – inherently ambiguous
  – complex and subtle
  – fuzzy, probabilistic
  – interpretation involves combining evidence
  – involves reasoning about the world
  – embedded a social system of people interacting
    • persuading, insulting and amusing them
    • changing over time

• Requires a process and training
Big Data Decision Method

1. Select Project
2. Define Questions & Criteria
3. Create Dictionary of key words
4. Create Decision Rules
5. Identify Sources
6. Big Data Work
7. Assess & Score Data on Criteria
8. Make Decisions

Iterative process
Examples

• Business Intelligence - PRA
• Customer Finder – Air Products
• New Business – Kelly Services
• Product Finder – Chemical Company
• Strategy Decisions – PentAir
It's the question; not the technology…

Transforming Clinical Trials through...
Our People, Innovation, Transparency
PRA International Approach

• Analyzed ~4 million data files totaling ~75 million web pages
• Linked the ‘big data’ with data from both ClinicalTrials.gov, and company-specific websites to answer the questions posed
PRA International Results

Questions
1. Which companies and institutions are conducting multiple myeloma clinical trials?
2. Who are the industry leaders and investigators involved?
3. Which clinical trials have failed and why?
4. What pharmacological products are currently being tested and which biochemical pathways are being targeted?
5. Are any companies working with these compounds or targeting these biochemical pathways, but not investigating multiple myeloma?

Answers
• 21 companies conducting 1249 trials in 54 sites
• Identified the managers of seven of these projects
• 91% were terminated/withdrawn due to lack of enrollment; 5% due to lack of compound, and 4% for undisclosed or other reasons
• Identified the 20 gene targets being used by these 21 companies
• Identified 7 companies using these same gene targets for indications e.g. asthma, breast and lung cancer, Parkinson’s, Alzheimer’s, sickle cell anemia, blindness
It's the question; not the technology...

Air Products’ Challenge
So Many Markets and Applications...
Use case

• Key questions:
  – Companies using industrial gases – which applications?
  – **Companies expanding or building new facilities?**
  – Companies who could be using gases but aren’t?
  – Companies we should know about but don’t?

• Limitations of SIC-based lists:
  – SIC’s are relatively broad and include too many non-industrial gas users
  – Segment specific detail to prioritize and focus lists not readily available
  – SIC’s may be incorrectly classified – missing relevant companies
The business value behind Natural Language Processing (NLP)

Search: Connect all words that trigger synonyms for capital spending with a cardinal number text string representing a specific dollar amount.

Outcome: Returns all text string related to any company increasing spending within the Metals Processing industry.

Investment spend

Capital spending programs totaling $61 million
Expanding in Arcadia, Louisiana and Kokomo, Indiana

As previously announced, the Company initiated capital spending programs totaling approximately $51.0 million to expand capacity at its Arcadia, Louisiana and Kokomo, Indiana facilities.

The Company spent $4.4 million on capital projects in the third quarter of fiscal 2012, which brings year-to-date capital spending to $17.1 million.
The forecasts for capital spending in fiscal 2013 and fiscal 2014 are approximately $70 million and $39 million respectively. The $70 million of capital spending planned for fiscal 2013 includes $19 million for the tubular project, $19 million for the Kokomo flat product project, $10 million for the processing and service center upgrades, $4 million for the information systems upgrade project and the remaining $16 million for additional enhancements and upgrades of the current facilities and equipment. The $39 million of capital spending anticipated for fiscal 2014 is expected to include $14 million to complete the Arcadia tubular project, $2 million to complete the Kokomo flat product project, $7 million to complete the processing center and service center upgrade, $2 million to complete the information system upgrade and the remaining $14 million for the continuing upgrade of the current equipment and facility. See Item 7. Management's Discussion and Analysis of Financial Condition and Results of Operations, section entitled Summary of Capital Spending in the Company Annual Report on Form 10-K for the fiscal year ended September 30, 2012 for additional discussion of individual projects.
It’s the question; not the technology…

Generate actionable business insight

Now the sales force for the industrial gas company knows:

1. **What** company is expanding – Haynes International, Inc.
2. **Where** the company is expanding
3. **When** they are spending money
4. **How** much money they spending
5. **Which** metals process is being invested in; therefore, which industrial gas product they are looking to purchase.

Seamless and welded pipe and tubing are manufactured in Arcadia, Louisiana. The smaller Arcadia facility consists of 125,000 square feet of factory on 47 acres.

18 months before construction began
Advancing Careers
Solving Workforce Challenges

We'll help you simplify your tasks, focus on your goals, and achieve success. That's because we're a leading workforce solutions provider that brings the best employees and employers together.
Project Description

Will the Flexible Distributed Workforce Cloud be viable in Healthcare?

How can Big Data be used to answer this question?
8-Step Big Data Decision Method

1. Select Project
2. Define Questions & Criteria
3. Create Dictionary of key words
4. Create Decision Rules
5. Identify Sources
6. Big Data Work
7. Assess & Score Data on Criteria
8. Make Decisions

Iterative process
1. Select Project: Intentions

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What problem/solution are you exploring?</td>
<td>If new service is a viable business</td>
</tr>
<tr>
<td>What do you want to know?</td>
<td>Client and employee joint needs</td>
</tr>
<tr>
<td>What market are you addressing?</td>
<td></td>
</tr>
<tr>
<td>What needs do you want to explore?</td>
<td></td>
</tr>
<tr>
<td>What technical area are you working to exploit?</td>
<td></td>
</tr>
<tr>
<td>What new sources of solutions do you want to explore?</td>
<td></td>
</tr>
<tr>
<td>What are you developing - a good or a service?</td>
<td>Service</td>
</tr>
<tr>
<td>What is the expected scope of the budget?</td>
<td></td>
</tr>
<tr>
<td>What resources are available?</td>
<td>Existing staff</td>
</tr>
<tr>
<td>What infrastructure is required?</td>
<td></td>
</tr>
<tr>
<td>What is the expected time frame?</td>
<td>18 months</td>
</tr>
<tr>
<td>What do you expect to return to the organization?</td>
<td>Higher margin revenue</td>
</tr>
</tbody>
</table>
2. Define Question Worksheet

Project Question: Is the FDWC a viable offering to the healthcare market?

(e.g., Is ____ (Product Y) ____ a viable offering to ____ (Customer X) ____)

<table>
<thead>
<tr>
<th>Consider the follow questions to form your project question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do you want to know? (strategic fit, demand, supply, competitors, environment, finance)</td>
<td>Will Kelly be able to enter the healthcare market with the FDWC?</td>
</tr>
<tr>
<td>What problem or opportunity do you want to explore?</td>
<td>What careers or areas within healthcare are best suited for the FDWC? How much demand and supply exists for the FDWC within healthcare?</td>
</tr>
<tr>
<td>What customer needs do you want to serve?</td>
<td>Are user groups and customer segments turning towards telehealth and telemedicine as an alternative to physical locations providing healthcare?</td>
</tr>
<tr>
<td>What capabilities do you want to test with the market?</td>
<td>Will the FDWC be able to both source talent and provide talent to meet the demand within the healthcare market?</td>
</tr>
<tr>
<td>What new markets do you want to explore?</td>
<td>Psychology, Physicians, Nurses of all kinds, Medical billers and coders, Medical transcriptionists</td>
</tr>
<tr>
<td>What new sources of solutions do you want to explore?</td>
<td>What is the supply of technically savvy healthcare professionals who are able to work remotely?</td>
</tr>
<tr>
<td>What is your revenue model?</td>
<td>“Kelly Specific”</td>
</tr>
</tbody>
</table>
# 3. Key Words (Dictionary) Worksheet

**Project Question:** Is the FDWC a viable offering to the healthcare market?

<table>
<thead>
<tr>
<th>Theme</th>
<th>Terms</th>
<th>Dictionaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Flexible/Distributed work (Virtual and/or Flexible jobs)</td>
<td>Work at home, flexible work, telecommute, etc.</td>
<td>Virtual Work Places to Work Virtually...</td>
</tr>
<tr>
<td>2. Work Force Cloud (Talent groups in telehealth)</td>
<td>Telemedicine, telestroke, remote care, telehealth, etc.</td>
<td>Freelance Job Portals Careers for Virtual Work</td>
</tr>
<tr>
<td>3. Healthcare</td>
<td>“portal”</td>
<td>Specific Healthcare Careers Healthcare Fields Telehealth...</td>
</tr>
<tr>
<td>4. Etc... “Market Landscape”</td>
<td>“portal”</td>
<td>Top 50 Pharma Top 10 Competitors, etc.</td>
</tr>
</tbody>
</table>
4. Rules Designer

- **Project Question**
- **Themes**
- **Terms**
- **Dictionaries**

Rule

1. Select
2. Define
3. Key word
4. Rules
5. Sources
6. Big Data
7. Assess
8. Decide
## 5. Source Worksheet

<table>
<thead>
<tr>
<th>Themes</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Flexible/Distributed work (Virtual and/or Flexible jobs)</td>
<td>Blogs, Associations, Freelance Job Portals, HHS.gov&lt;br&gt;SHRM, Telework Research Network, Public periodicals</td>
</tr>
<tr>
<td>2. Work Force Cloud (Talent groups in telehealth)</td>
<td>Freelance Job Portals, Work related blogs&lt;br&gt;“oDesk, WAHM”</td>
</tr>
<tr>
<td>3. Healthcare</td>
<td>ATA, OAT, CTeL, BLS, HRSA, TRCs, APA, Hospital websites&lt;br&gt;Insurance websites, Technology company websites, etc.</td>
</tr>
<tr>
<td>4. Etc… “Market Landscape”</td>
<td>Top 50 Pharma companies, Top 10 competitors, Job websites&lt;br&gt;University Websites (Miami, Utah, California, Virginia, etc.)</td>
</tr>
</tbody>
</table>
6. Big Data Work: NLP stack

Write Rules → Process Data → Manage Data

Data Decision Modeler → Analyze Results → Data Scientist

It's the question; not the technology...
7. Data Assessment Scorecard

<table>
<thead>
<tr>
<th>Data Categories</th>
<th>Data Type</th>
<th>Relevance (1–5)</th>
<th>Amount (1–5)</th>
<th>Score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Demand</td>
<td>A, G, N, B</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Supply</td>
<td>G, SM, N, B</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Competition</td>
<td>B, N</td>
<td>4</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Environment</td>
<td>A, G, SM, N, B</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Finance</td>
<td>A, G, N</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>20</strong></td>
<td><strong>18</strong></td>
<td><strong>38 of 50</strong></td>
</tr>
</tbody>
</table>

*Decision Ready (average score by “Data Category”) (Can add new “Data Types”)
## 7. Results Score Card

<table>
<thead>
<tr>
<th>Decision Criteria</th>
<th>SME Ratings: 1=low, 5=high</th>
<th>Score</th>
<th>Critical insight from the results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategic fit</td>
<td>1=no fit 5=high fit</td>
<td>2</td>
<td>“Kelly Specific”</td>
</tr>
<tr>
<td>Demand</td>
<td>1=little 5=high</td>
<td>5</td>
<td>Decreased healthcare costs, increased quality Academic, peer reviewed, proof</td>
</tr>
<tr>
<td>Supply</td>
<td>1= not in Kelly 5= In place</td>
<td>2</td>
<td>Not met by regional providers A handful of accredited training institutions</td>
</tr>
<tr>
<td>Competition</td>
<td>1= established 5=non-existent</td>
<td>3</td>
<td>Segmented by region, larger players focus on telehealth technologies, follow reimbursements</td>
</tr>
<tr>
<td>Environment</td>
<td>1=Hostile 5=Receptive</td>
<td>3</td>
<td>High segmentation, Highly regulated Defined policies, easier to predict changes</td>
</tr>
<tr>
<td>Finance</td>
<td>1=little potential 5=High potential</td>
<td>4</td>
<td>Focus on high value talent</td>
</tr>
</tbody>
</table>
It's the question; not the technology…
8. Decide: Project Results

Will the Flexible Distributed Workforce Cloud be viable in Healthcare? Yes

- **Drivers:**
  - Access to healthcare
  - State and Federal policies
  - Technology enablement

- **Success Factors:**
  - Market segmented based on reimbursement and geography,
  - Partnerships with regional telemedicine centers
  - Market positioning is critical

- **Business Model:**
  - Multi-sided business model
Building Business Models with Big Data

- Move ideas from political to analytical
- Cross the Valley of Death
- Structured process for building business models
- Substantiate business models with big data
- Pick the right tool – un/structured
- Know your questions
It's the question; not the technology…

END
“Organizations will not only need to ensure they have sufficient skills in back office analytics but also manage a transition toward the right managerial talent on the front line that will be capable of executing strategy based on the insights analysts mined from big data.”
INTRODUCTION

- Benefits
  - People focused on high value projects
  - Needs are anticipated & addressed
  - Project reviews – kill or continue
  - People are “authorized” to work on innovation projects
  - Standardized process reduces time
  - Holistic view of long-term research
  - Reputation enhancement

More ideas screened in less time at a lower cost with higher value
OPPORTUNITY FRAMEWORK

• Worksheets
  – Intension
  – Idea
  – Elaborate
  – Difference and Benefits
  – Ability to Deliver
  – Opportunity Statement

It’s the question; not the technology…
## Opportunity Statement

<table>
<thead>
<tr>
<th>State Your Idea:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe how the idea is grounded in an enduring customer need</td>
</tr>
<tr>
<td>Describe your existing or new ability that will deliver on the need.</td>
</tr>
</tbody>
</table>
### Key Partners
- Who are our Key Partners?
- Who are our key suppliers?
- Which Key Resources are we acquiring from partners?
- Which Key Activities do partners perform?

#### Motivations for partnerships:
- Optimization and economy
- Reduction of risk and uncertainty
- Acquisition of particular resources and activities

### Key Activities
- What Key Activities do our Value Propositions require?
- Our Distribution Channels?
- Customer Relationships?
- Revenue streams?

#### Categories
- Production
- Problem Solving
- Platform/Network

### Value Proposition
- What value do we deliver to the customer?
- Which one of our customer’s problems are we helping to solve?
- What bundles of products and services are we offering to each Customer Segment?
- Which Key Resources are we acquiring from partners?

#### Categories
- Newness
- Performance
- Customization
- “Getting the Job Done”
- Design
- Brand/Status
- Price
- Cost Reduction
- Risk Reduction
- Accessibility
- Convenience/Usability

### Customer Relationships
- What type of relationship does each of our Customer Segments expect us to establish and maintain with them?
- Which ones have we established?
- How are they integrated with the rest of our business model?

#### Examples
- Personal assistance
- Dedicated Personal Assistance
- Self-Service * Automated Services
- Communities * Co-creation

### Key Resources
- What Key Resources do our Value Propositions require?
- Our Distribution Channels?
- Customer Relationships?
- Revenue Streams?

#### Types of resources
- Physical
- Intellectual (brand patents, copyrights, data)
- Human
- Financial

### Channels
- Through which Channels do our Customer Segments want to be reached?
- How are we reaching them now? How are our Channels integrated?
- Which ones work best?
- Which ones are most cost-efficient?
- How are we integrating them with customer routines?

#### Channel phases:
1. Awareness
2. Evaluation
3. Purchase
4. Delivery
5. After sales

### Customer Segments
- For whom are we creating value?
- Who are our most important customers?

#### Market Types
- Mass Market
- Niche Market
- Segmented
- Diversified Multi-sided Platform

### Cost Structure
- What are the most important costs inherent in our business model?
- Which Key Resources are most expensive?
- Which Key Activities are most expensive?

#### Is your business more:
- Cost Driven (leanest cost structure, low price value proposition, maximum automation, extensive outsourcing)
- Value Driven (focused on value creation, premium value proposition)

#### Sample characteristics:
- Fixed Costs (salaries, rents, utilities)
- Variable costs
- Economies of scale
- Economies of scope

### Revenue Streams
- For what value are our customers really willing to pay?
- For what do they currently pay?
- How are they currently paying?
- How much does each Revenue Stream contribute to overall revenues?

#### Types:
- Asset sale
- Usage fee
- Subscription Fees
- Lending/Renting/Leasing
- Licensing
- Brokerage fees
- Advertising
Why this is **not** Google

<table>
<thead>
<tr>
<th>Google</th>
<th>Big Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key <em>word</em> search</td>
<td>Multiple word lists</td>
</tr>
<tr>
<td>Limited rules - Proximity</td>
<td>Natural language: Nouns, Verbs, Sentences, Paragraphs,</td>
</tr>
<tr>
<td>Unlimited results</td>
<td>Targeted results</td>
</tr>
<tr>
<td>Biased results (page ranking battles)</td>
<td>Unbiased results</td>
</tr>
</tbody>
</table>
Big Data Workflow

Sources of Information → Gather/Crawl → Raw Information File → Index → Data File → Analyze → Knowledge

It’s the question; not the technology…
Framework supporting **Map-Reduce** Programming Paradigm

It’s the question; not the technology…
It’s the question; not the technology…
It’s the question; not the technology…
Big Data training for is the key for meeting needs and increasing sales
Air Products Chinese Project

Shenyi Zuo
Aug 13, 2013
Agenda

• Project Introduction
• Project Timeline
• Dictionaries & Rules
• Key Learning from Chinese Language & ICA
• Q & A
Project Introduction

Metal Fabrication & Processing

- Aerospace
- Automotive
- Shipbuilding

- Auto part suppliers’ addresses
- Tier 1 auto companies’ investment
- Trends of new energy automotive
Project Timeline

Jun
- Receive project brief
- Build 1st iteration dictionary & model
- Learn the market

Jul
- Narrow down the scale
- Improve the dictionary, source, and model with Shanghai team
- ICA training

Aug
- Results
- Presentation
- Final report
# Decision Model

<table>
<thead>
<tr>
<th>Step</th>
<th>Step</th>
<th>Example</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Project Selection</td>
<td>Air Products Metals Processing and Fabrication in China</td>
<td>Project Worksheet</td>
</tr>
<tr>
<td>2</td>
<td>Project Question</td>
<td>Potential customers in China auto part supplier industry: company list &amp; contacts, buyer relationships</td>
<td>Project Question Worksheet</td>
</tr>
<tr>
<td>3</td>
<td>Create Dictionaries</td>
<td>company, political environment, investment, partnership, fabrication &amp; processing terms, processing gas, auto manufacture terms</td>
<td>Dictionary Worksheet</td>
</tr>
<tr>
<td>4</td>
<td>Identify Sources.</td>
<td>Metals processing and fabrication industry key players’ websites, auto-part supplier websites, government websites</td>
<td>Source Worksheet</td>
</tr>
<tr>
<td>6</td>
<td>Big Data Workflow</td>
<td>Raw Data File, Meta Data File/Map reduce, Natural Language Indexed data file. Map reduce.</td>
<td>Results file: Tagged Information</td>
</tr>
<tr>
<td>7</td>
<td>Data Assessment</td>
<td>Quantity, quality, relevance</td>
<td>Data Assessment Worksheet</td>
</tr>
<tr>
<td>8</td>
<td>Results Score Card</td>
<td>Prefilled score card</td>
<td>Trends and Insights, AHP Score Card, Data Summary</td>
</tr>
</tbody>
</table>
## Project Question Worksheet

<table>
<thead>
<tr>
<th>Consider the follow questions</th>
<th>Specific questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do you want to know?</td>
<td>Addresses of potential customers in China auto part supplier: company list, partnerships/contacts, buyer relationship, locations, and etc.</td>
</tr>
<tr>
<td>(strategic fit, demand, supply, competitors, environment, finance)</td>
<td></td>
</tr>
<tr>
<td>What problem or opportunity do you want to explore?</td>
<td>What’s the current supplier market map in China? Are new regions being developed? What are the market trends in China?</td>
</tr>
<tr>
<td>What new markets do you want to explore?</td>
<td>How is the new energy market growing?</td>
</tr>
<tr>
<td>What new sources of solutions do you want to explore?</td>
<td>What industrial parks will survive industry downsizing? What are the main industrial regions?</td>
</tr>
</tbody>
</table>
## Dictionary Worksheet

<table>
<thead>
<tr>
<th>Theme</th>
<th>Terms</th>
<th>Dictionaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Automotive market</td>
<td>Auto company, local auto manufacture, auto engine manufacture term,</td>
<td>Market</td>
</tr>
<tr>
<td>2 Facts that effects markets</td>
<td>Partnership, political environment, business, investment, general region</td>
<td>Commercial Environment</td>
</tr>
<tr>
<td>3 Manufacture process</td>
<td>Fabrication, processing, processing gas</td>
<td>Technology</td>
</tr>
</tbody>
</table>
# Source Worksheet

<table>
<thead>
<tr>
<th>Themes</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Auto part supplier company</td>
<td>Industry association websites</td>
</tr>
<tr>
<td></td>
<td>B2B websites</td>
</tr>
<tr>
<td></td>
<td>Key players’ websites</td>
</tr>
<tr>
<td>2 Policy</td>
<td>National government websites</td>
</tr>
<tr>
<td></td>
<td>Local government department</td>
</tr>
<tr>
<td>3 Technology</td>
<td>Online industrial magazine</td>
</tr>
</tbody>
</table>
## Rules

<table>
<thead>
<tr>
<th>Themes</th>
<th>Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto supplier company</td>
<td>Auto investment: [data]X[auto company]X[general region]X[investment]</td>
</tr>
<tr>
<td></td>
<td>Local investment: [local auto company]X[general region]X[investment]</td>
</tr>
<tr>
<td></td>
<td>Supplier finder: [company finder]X[fabrication]</td>
</tr>
<tr>
<td></td>
<td>Supplier add finder: [supplier finder]x[add]</td>
</tr>
<tr>
<td>Trends</td>
<td>Market trend: [general region]X[investment]X[new energy]</td>
</tr>
</tbody>
</table>
Key Learning

• Improve the model through testing
• Building rules under the Chinese grammar structure. Difference: Chinese doesn’t have times, so how to define a noun & verb?
• Building dictionaries not only by categories, but also by attributes, because the ICA cannot identify a Chinese sentence.
It’s the question; not the technology…

[Auto company]x[general region]x[investment]