

BigSheets

Addressing the challenge of big data for
business insight and analysis



solutions start here.

What does BigSheets Address?



The Challenge: Companies are awash in “big data”.

Significant amounts of data, and no simple, easy way for companies to discern business insight and identify opportunities hidden within that data. This problem is only growing, by some estimates 988 exabytes of information will be generated in 2010. That's over 18 million times the amount of information contained in every book ever written over the course of human history.

- Volume of data across enterprises is doubling approximately every three years².
- Web data management—not just database management—will counter this data escalation.
- Web content—and the people behind it—is an untapped source for business intelligence.
- Costs to find, collect, and analyze data are projected to decrease significantly as the web matures—but the tooling to accomplish this task are lacking...**which is where BigSheets steps in...**

¹ Source: IDC, "An Updated Forecast of Worldwide Information Growth Through 2011", March 2008

² Source: Forrester Research (Oct 17, 2008)



The Solution: BigSheets



What is BigSheets? A revolutionary browser-based analytics tool that enables business users to extend the scope of their intelligence data through the web in a timely manner.

Driving Factors:

- Business users require a **non-technical** approach to break down massive amounts of data into consumable, situation-specific business contexts.
- Lucrative opportunities exist for translating untapped, unstructured (and often unknown) data into actionable business insights.
- Success will be measured by entities that can slice and dice both enterprise and web data into consumable, re-mixable business intelligence.

How BigSheets Helps:

- BigSheets has a UI developed specifically for business intelligence and non-technical business users to facilitate data gathering and analysis.
- BigSheets has the ability to gather data in structured form (databases, etc.) and unstructured form (web sites, documents, etc.)
- By allowing the combination of these two data sources, BigSheets allows business users ID opportunities and risks “hidden in the data”.



Scenarios & Business Patterns:

- **Financial Services** want to improve the scope of IP portfolio due diligence analyses for M&A.
- **Healthcare/Life Sciences Companies** want to identify, meter, and monitor top drug prescribers. Both payers and providers need to understand what their data is telling them.
- **Retail Companies** need to understand their customers better and discover ways to optimize operations given the current economic challenges.
- **Telecommunications Providers** mine opinion websites and blog comments to discover customer attitudes about cell phone & service plan offerings.
- **Media & Entertainment Companies** want to quickly discover and expose exclusive, novel content to attract more viewers & advertisers.

BigSheets in Action:

- **Gather** gigabytes, terabytes—even petabytes—of unstructured web data.
- **Enrich** and **Inspect** that data to better correlate with your business search.
- **Publish** the results so that more business users can begin **Exploring** them.



A Sample Scenario



An international new media company...

wants to gain additional advertising business through an innovative, novel content platform focusing on UK Parliament—but researching structured news databases combined with Google searching may not be the best way to perform news research...

The Need:

To leverage the broader internet for all relevant content...

Previously Inaccessible Data:

- Cross Divisional
- Ecosystem
- User generated
- News Feeds
- MySpace
- Facebook
- Twitter
- etc...

BigSheets Inputs...

1. UK Parliament Site & related demographic / voting records.
2. Members of Parliament & related Bills, Debates, Districts, etc...
3. Time (10+ years)

...to find out...

- Who is doing what – and who isn't doing?
- Overlay voting records with demographics and voting records over time



Gathering, Inspecting, & Enriching



Gather collections of data from multiple web sources:
Entities, Bills, Members of Parliament, & Parties.

Inspect the entities collection data for consistency & relevance by **Type, Count, Rank & Value.**

M2 Welcome Joe the LOB | Settings | Logout

Search collections... search

Collections Recent Favorites **Crawlers** New Monitor

Open Calais Entities
 Open Calais returns a collection of people, places, things and industry terms based on the analysis of the Web Service hosted by Reuters
 Collecting... January 22, 2009

Bill Pages
 The BillPages collection represents the Bills in the UK Parliament that have been debated, passed or dismissed over the past 10+ years
 Collecting... January 22, 2009

Members of Parliament
 This collection represents the current members of the UK Parliament and the constituency supported by the MP
 Collecting... January 22, 2009

Members of Parliament & Party
 This collection represents the current members of the UK Parliament and their associated party
 Collecting... January 22, 2009

UKParliament <status>

1 Item 12/12/08 12:20:11AM	80 Items 12/12/08 12:22:32AM	92,1234 Items 12/12/08 07:10:45AM
----------------------------------	------------------------------------	---

Grid Graph Chart

Value	Url
Iraq	http://www.publications.parliament/.../357/index-25.htm
	http://www.publications.parliament/...

Enrich your data set by creating relationships between collection data:
BillName, PersonName, Count, & Rank

Calculate:

TotalCount = SUM(Count) Fx + -

AverageRank = AVG(Rank) Fx + -

BillName	PersonName	TotalCount	AverageRank
Animal Welfare	B. Bledisloe, V	1	0.123
Animal Welfare	B. Judd	3	0.426
Animal Welfare	Ben Bradshaw	7	0.47
Animal Welfare	L. Bridgeman, V	4	0.503
Animal Welfare	Simon, V	7	0.39
Crossrail	Bridget Prentice	1	0.22



Publishing & Exploring



Open Calais Entities

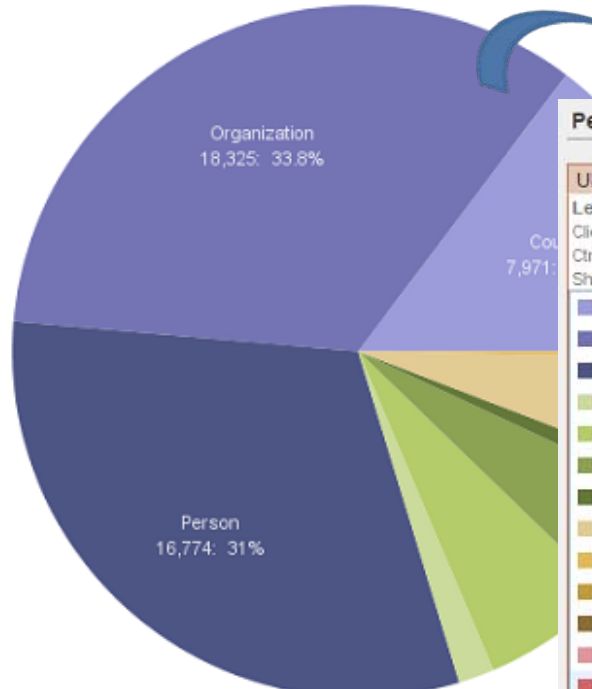
UK Bill & Debate Entities

Legend

Click to select,
Ctrl-Click: multiple
Shift-Click: range

Count (aggregate). Total=54,177

Country	7,971
Organization	18,325
Person	16,774
ProvinceOrState	974
City	3,377
IndustryTerm	2,969
Technology	520
Company	3,029
Event	238



Publish massive results to begin sorting entity collection types by **UK Bill & Debate count.**

People

UK Parliament People Associated with Parliament

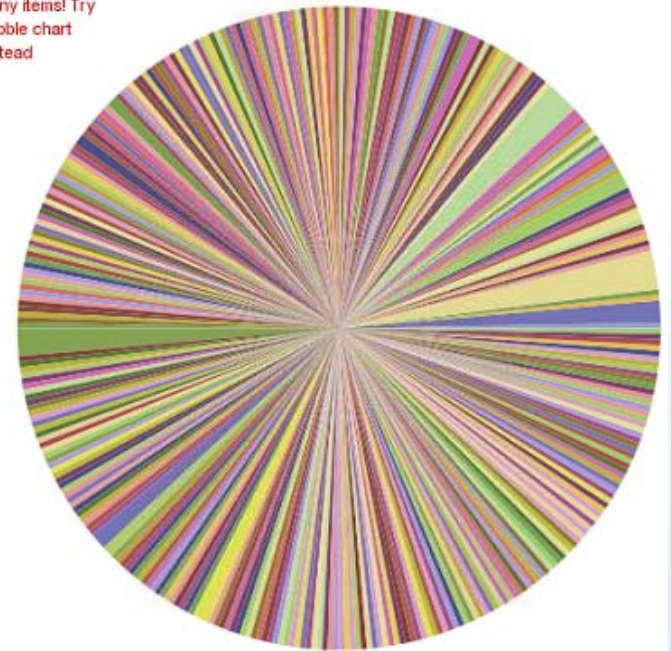
Legend

Click to select,
Ctrl-Click: multiple
Shift-Click: range

Rank (aggregate). Total=16,680

Many items! Try
bubble chart
instead

Boswell	16
Lea	173
Mudie	1
Advocacy	1
Alan Haselhurst	7
Alexander	29
Andrew (Westbury)	1
Angela	3
Children	3
Clive	3
Elizabeth Filkin	1
Izett	1
Mike Hall	1
Rosie	2
Royal Naval Reserve	1
Betty (Conwy)	2
Colin Powell	7
David Curry	3
George	14



Explore entity type person to view ranking of most active people. **Who is doing what—and who isn't doing?**



Business Value Revealed



Problems

Research process is slow and costly.

Advertising revenue may be lost to faster competitors with more exclusive content.

Solution

Develop a new media platform leveraging web data (via M2) to drive consumer and advertising traffic by providing UK Parliament transparency.

Value

Research productivity improves as M2 UK Parliament data collections grow.

Web data mined by M2 provides novel UK Parliament news.



Sounds promising, so where do I begin?

jStart can help you understand, evaluate, and even prototype possible use of the BigSheets technology for your company. The team has vast experience working with clients to understand unique needs and challenges facing enterprises, and how new and novel technologies can be leveraged to address them. To get started, simply choose the most convenient method to contact the team...

Contacting & Learning about jStart

Visit our website:

www.ibm.com/jstart

Contact us via the website

www.ibm.com/software/jstart/contact.shtml

Contact us via email:

jstart@us.ibm.com

Working with jStart

jStart has a model which it engages customers/clients with a new technology...see which one best fits your needs...

- 1. Briefing.** This includes presentations, roundtable discussions, off the shelf demonstrations and high level architectural white board sessions exploring the challenges the customer faces and how BigSheets can address those challenges.
- 2. QuickStart Pilot.** Provides an advanced BigSheets solution for analyzing big data. The options is best suited for customers who are ready to start a small pilot project.
- 3. Integration Workshop.** Need a more comprehensive solution to satisfy your business needs? jStart can provide a solution with capabilities for specific domains, applications, and customers.



Thanks!

Contact us to discover how BigSheets
can give your business a competitive edge.

www.ibm.com/jstart/



solutions start here.