Overview

- Blackbook2 is a J2EE server-based data integration framework

- Relies on open standards to promote robustness and interoperability
  - JENA, JUNG, Lucene, JAAS, D2RQ

- Based on semantic web technologies
  - RDF, RDF Schema, OWL, SPARQL
  - Vocabulary agnostic

- Provides a default web application interface, SOAP and RESTful interfaces

- Blackbook2 is PL3 Appendix E certified (PL3+)
Architecture

Visualization Layer

Infrastructure Layer

Data Source Layer
A front-end "Google-like" user interface allows analysts to easily perform keyword and attribute based searches.
User Interface

"Google-like` Results

Network

Different ways to view the same information. “Network”, for example, displays entities of different types and their relationships to other entities.
User Interface

Different ways to view the same information. “Timeline”, for example, displays entities chronologically.
User Interface

Google Map

Google Earth

Allows analysts to visualize geospatial content using Google-map and Google Earth.
User Interface

Analyst Notebook

Mediawiki
Architecture

Visualization Layer

Infrastructure Layer

Data Source Layer
The Semantic Web is the next generation of the current web in which computers can interpret the meaning of the web content because of explicit semantics provided in markup.
Example 1: Inference

An analyst creates:
1) Entity "Buster"
2) Entity "Jennifer"
3) Entity "Nicholas"

An analyst makes the assertion:
4) "Buster husband-of Jennifer"
5) "Jennifer mother-of Nicholas"

Blackbook system can infer:
6) "Jennifer wife-of Buster"
7) "Nicholas child-of Jennifer"
8) "Buster [step]father-of Nicholas"
9) "Nicholas [step]child-of Buster"
10) "Buster is-gender Male"
11) "Jennifer is-gender Female"
Example 2: Invalid Logic Assertion

An analyst creates:
1) Entity "Don"

An analyst makes the assertion:
2) "Jennifer brother-of Don"

Blackbook system can infer:
3) Invalid Assertion
   (Gender conflict)
Example 3: Constraints & same-as

An analyst makes the assertion:
1) "Buster has-brother Bill"
2) "Buster has-sister Candy"
3) "Candy has-son Mark"
4) "Mark has-uncle William"

An analyst applies the constraint:
5) "Buster has-only-one-brother Bill"

Blackbook system can infer:
6) "William same-as Bill"
7) "Bill same-as William"
Algorithm plug-ins can be added

Security PL3+ / User Credentials
- CASPORT
- Scattered Castles
- DIAS

Analysts can store assertions into an Analytic Knowledge Base (AKB)
"Workflow" allow analysts to define the order of tasks, configure algorithm parameters, and batch processes concurrently.
"Workflow" and "Workspace" allow analysts to define the order of tasks, store them in private folders and/or share them publicly with colleagues.
Architecture

Visualization Layer

Infrastructure Layer

Data Source Layer
Composite Knowledge

Original Datasource

1 =

2 =

3 =

Analyst Knowledge Base

Composite Knowledge

AKB
Composite Knowledge with Security

Original Datasource

Analyst Knowledge Base

Composite Knowledge

Unclassified
Secret
Top Secret
Composite Knowledge with Confidence

Original Datasource

Analyst Knowledge Base

Composite Knowledge

Original Datasource:
- U
- S
- T

Analyst Knowledge Base:
- U
- S
- T

Composite Knowledge:
- U
- S
- T

AKB (Analyst Knowledge Base)
User Interface

- **Relationship Manager**
  - Allows analysts to specify the relationship between two or more entities.

- **Entity Manager**
  - Allows analysts to create entities of different types, and modify attributes.

- **Ontology Import**
  - Allows analysts to upload their own ontology.
Unlike most applications, Blackbook performs queries on data in RDF form, not relational form.
Datasource Connectivity

Currently, there are six avenues to connect Blackbook with data; 1-3 requires offline translation to RDF, 4-5 uses "real-time" translation using D2RQ, 6 uses secure web-services.
Blackbook and Alternate Stores

Hbase, Lucene, Solr
Scalability using Hadoop

Hadoop implements MapReduce, using the Hadoop Distributed File System (HDFS). MapReduce divides applications into many small blocks of work. HDFS creates multiple replicas of data blocks for reliability, placing them on compute nodes around the cluster. MapReduce can then process the data where it is located.

**Scalable:** Hadoop can reliably store and process petabytes.  
**Economical:** It distributes the data and processing across clusters of commonly available computers. These clusters can number into the thousands of nodes.  
**Efficient:** By distributing the data, Hadoop can process it in parallel on the nodes where the data is located. This makes it extremely rapid.  
**Reliable:** Hadoop automatically maintains multiple copies of data and automatically redeployed computing tasks based on failures.
Blackbook and Wikis
Blackbook and Wikis

Like browsers, "Wiki" can be a front-end to Blackbook. Wikis can also be a datasource. Wiki extensions can be utilized to enable Semantic and Blackbook features.
A research product (red), such as a new and improved algorithm or visualization, can easily be transferred from research to government using the Blackbook “envelope”. 

Revolutionize Technology Transfer

Improve Intelligence Analysis by Coordinated Exposition of Multiple Data Sources Across Intelligence Community Agencies
Relational vs. Graph-based Systems

Most Systems

Relational

Data Source Layer

Infrastructure Layer

Graph

Data Source Layer

Infrastructure Layer

Blackbook2 is a JEE server-based RDF processor that provides an asynchronous interface to back-end datasources.