Recovering Business Rules from Legacy Source Code for System Modernization

Erik Putrycz, Ph.D.
Anatol W. Kark
Software Engineering Group
National Research Council, Canada

NRC·CNRC
Institute for Information Technology

Introduction

- **Legacy software**

```
000009* PROGRAM TITLE:
000010* DATE CREATED:  JULY 1979
000011* AUTHOR:        UNKNOWN
000012* DESCRIPTION:
```

- **Common**
  - In 2006, 70% of all transaction systems were written in COBOL

NRC·CNRC
Institute for Information Technology

Introduction (2)

- **Modernization**
  - Motivations
    - High cost to operate legacy system
    - Impossible to keep the legacy system up-to-date
    - Lack of qualified staff
  - New system or integrate legacy with new system
    - Need for requirements
  - Many requirements buried in the source code
  - Recovering business rules major issue
    - Recent survey from Software AG: 51% of companies who have difficulties modernizing said that a major issue are “hard-coded and closed business rules”
- **Objective**
  - How to recover business rules from legacy source code?

NRC·CNRC
Institute for Information Technology

Outline

- **Context**
- **Extraction process**
- **Visualization and navigation**
- **Conclusions**

NRC·CNRC
Institute for Information Technology

Outline

- **Context**
- **Extraction process**
- **Visualization and navigation**
- **Conclusions**

NRC·CNRC
Institute for Information Technology

Legacy Systems

- They handle and process a large amount of data.
- They have a large user base.
- They involve many business rules linked to legislation, agreements, processes, etc.
- Many business rules are not accurate anymore and require external error and exception processing because of the difficulty to improve and update such system.
- The maintenance of the system often involves changes related to business rules.
- The source code is available – most developments “in-house”.

NRC·CNRC
Institute for Information Technology

Stakeholders

- Two main classes of the stakeholders are involved with business rules:
  - *Legacy system maintainers*
    - Fix bugs and implement new rules.
    - Need to understand the business rules they are affecting and the execution paths to a specific business rule
  - *Business analysts*
    - Involved in modernization of the legacy system
    - Business rules in legacy system used for validating new requirements or finding requirements
    - Often no background in technologies used in legacy system

NRC·CNRC
Institute for Information Technology

Objectives

- **Build an extensible and customizable framework for knowledge extraction**
- **Extract business rules in the following form:**
  - If <conditions> then <consequence>
  - <conditions> and <consequence> as easy to understand as possible
- **Use "business terms" instead of programming language constructs**
- **Focus on calculations**
  - Representation of complex calculations in an "easy" format
- **Provide navigation and visualization to locate information**
- **Focus on COBOL legacy software**

NRC·CNRC
Institute for Information Technology

Outline

- Context and objectives
- **Extraction process**
- Visualization and navigation
- Conclusions

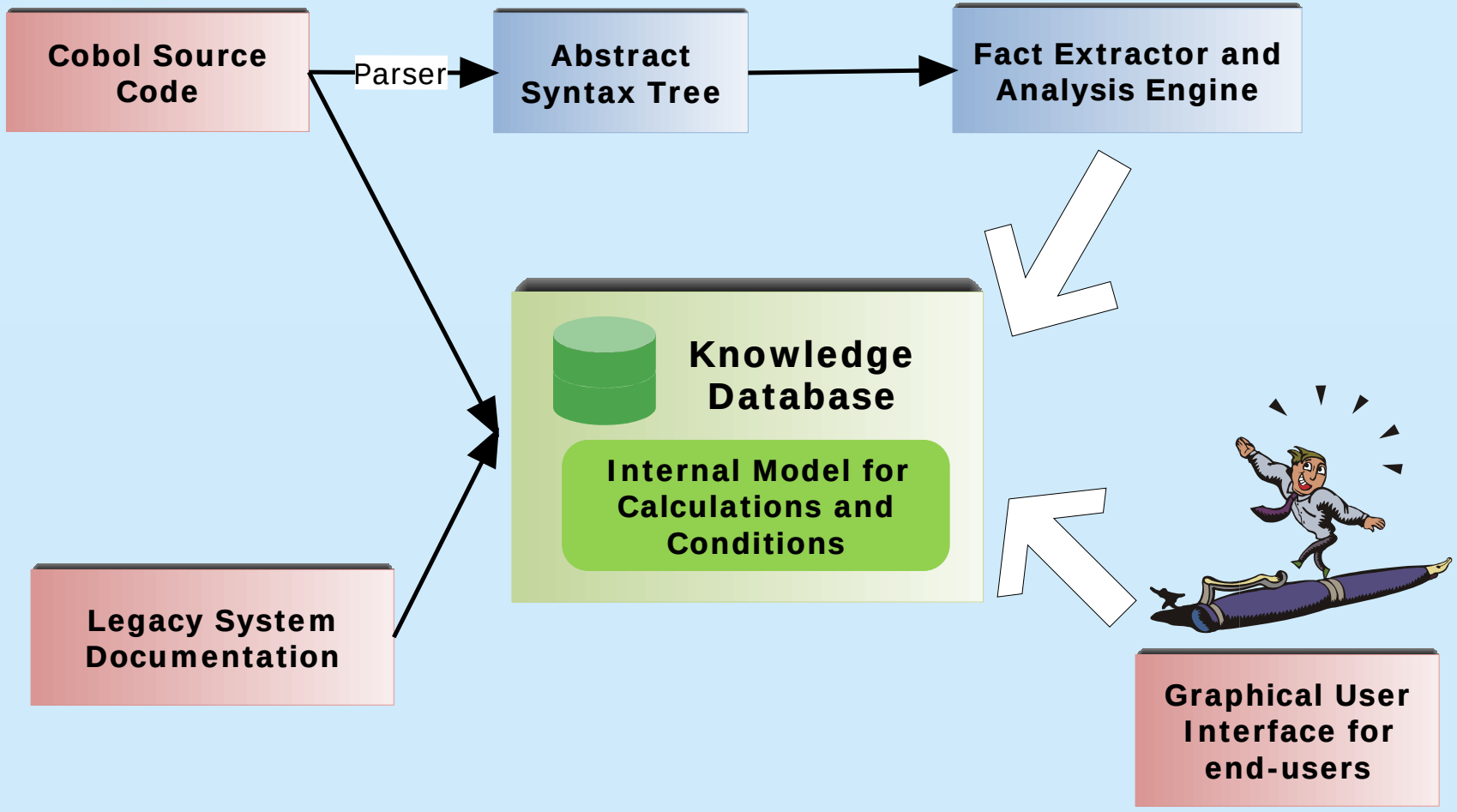
NRC·CNRC
Institute for Information Technology

COBOL

- **Limited and simple structure**
  - each program = sequence of statements grouped into paragraphs which are executed sequentially
- **Only two forms of branching**
  - one to execute external programs (CALL)
  - another to transfer the control flow to a paragraph (PERFORM). PERFORM is also used for iterations.
- **Each program is associated with a single source file.**
- **Elements of Business Rules**
  - Branching: execute a paragraph
  - Calculations: compute calculations between variables into another
  - Conditions

NRC·CNRC
Institute for Information Technology

Platform overview



NRC·CNRC
Institute for Information Technology

Knowledge database

- **Main elements**
  - Business rules = condition + action
  - Action: calculation or control flow branching
  - Conditions: Boolean operations between identifiers
  - Calculations: arithmetic operations between identifiers
  - Identifiers
- **Keyword indexing all of elements**
- **Source code**
  - All extracted artifacts are linked to source code
- **Variable names translation to business terms**

NRC·CNRC
Institute for Information Technology

From Source code to Business Rules

- **Abstract Syntax Tree**
  - Extract machine level representation of the source code
- **Business Rules construction**
  - Locate calculations
  - Locate conditions
    - Combine nested ifs
  - Simplify calculations
  - Simplify conditions
  - Compute dependencies between business rules
- **Integrating documentation and comments**
  - Attach comments to relevant elements

NRC·CNRC
Institute for Information Technology

Implementation

- **Technologies**
  - Object oriented database DB4O for storing and querying business rules
  - Antlr3 (LL-*) parsing for generating AST from COBOL
- **One major legacy system**
  - 630000 lines of COBOL
  - 18 programs
  - 11924 business rules
  - 9292 identifiers
  - 2823 paragraphs
- **Database generation**
  - 4h for generation for extracting all business rules and indexing

NRC·CNRC
Institute for Information Technology

Outline

- Context and objectives
- Extraction process
- **Visualization and navigation**
- Conclusions

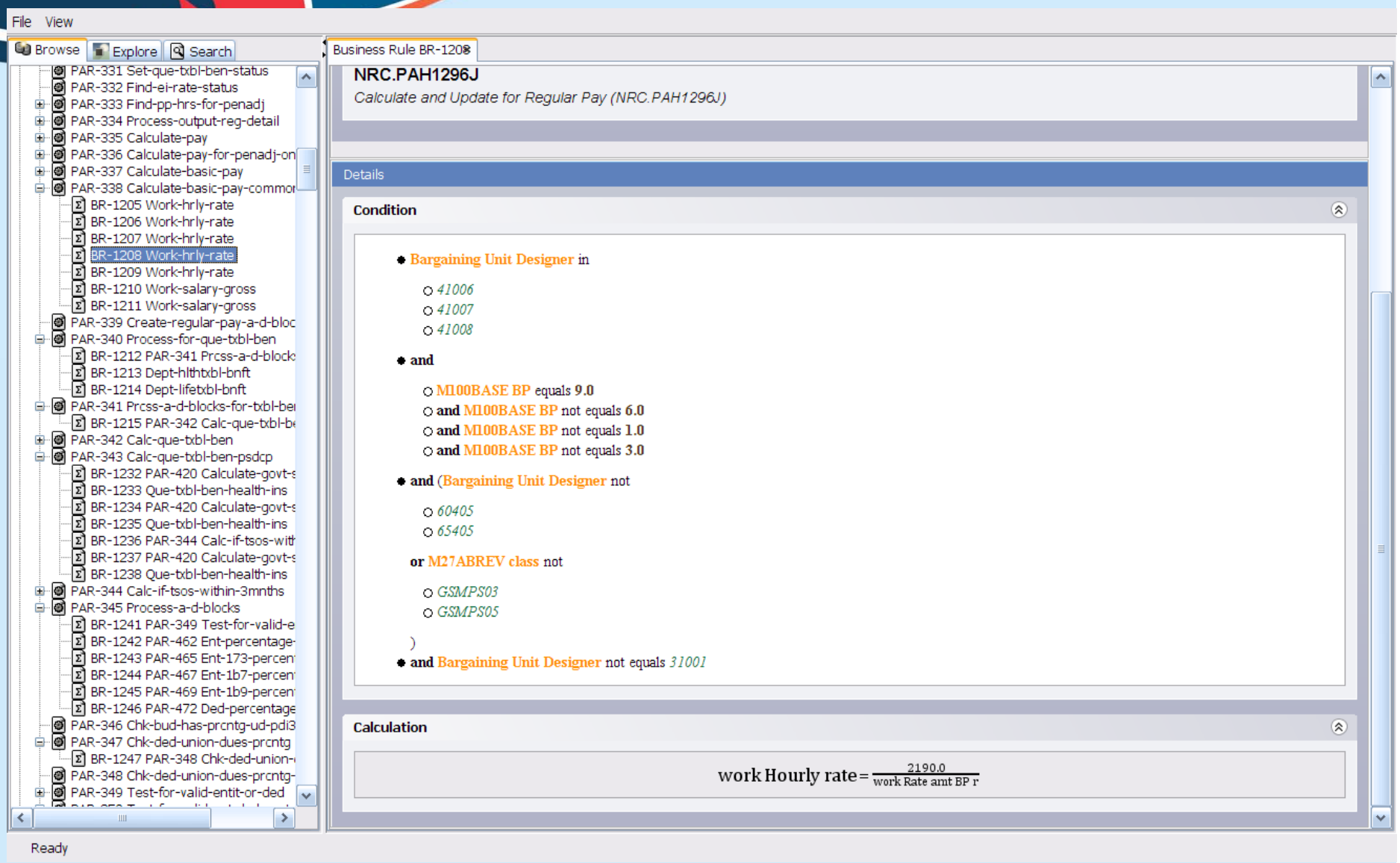
NRC·CNRC
Institute for Information Technology

Visualization

- **Large amount of data**
  - 18 programs
  - 11924 business rules
- **Business analysts**
  - Main artefacts are requirements
- **Need for locating relevant information to users**
  - System wide navigation
- **Recent studies [Ko and al. 2006]**
  - flexible navigation is a key element for software maintainers

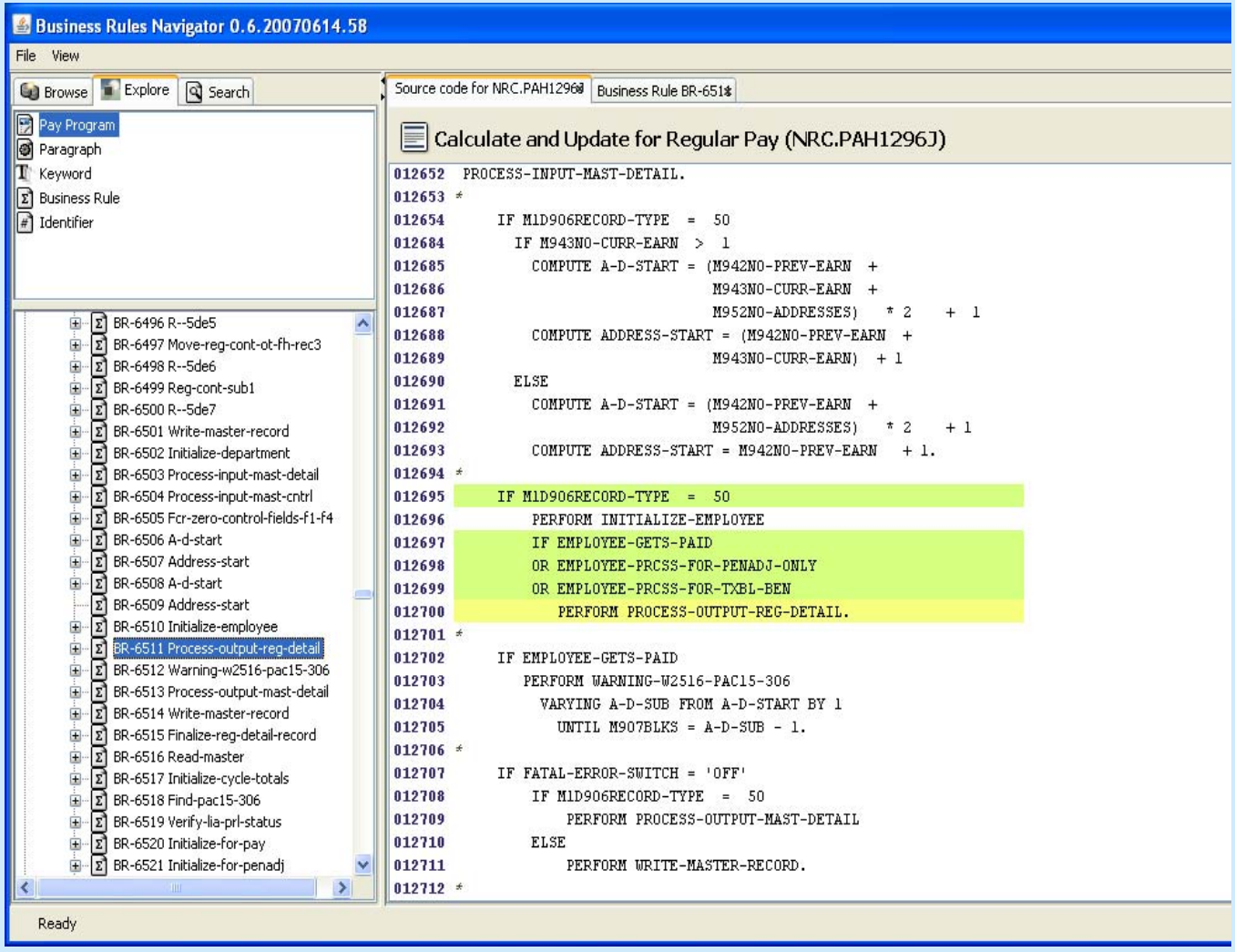
NRC·CNRC
Institute for Information Technology

Main User Interface



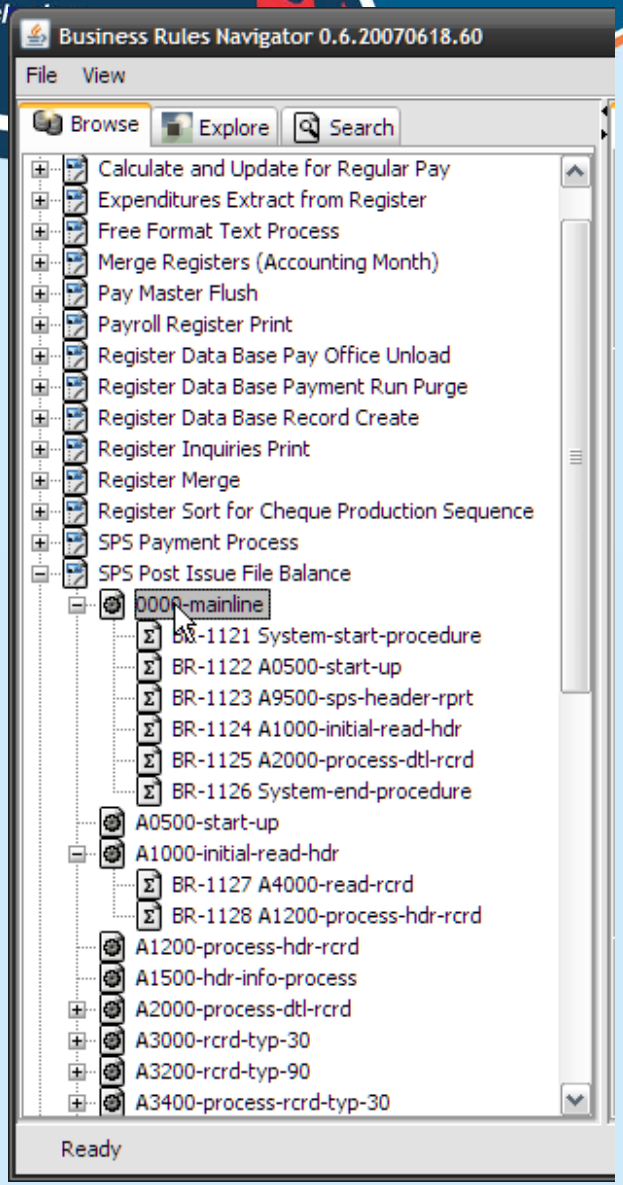
NRC·CNRC
Institute for Information Technology

Business rules and source code linking



NRC·CNRC
Institute for Information Tech

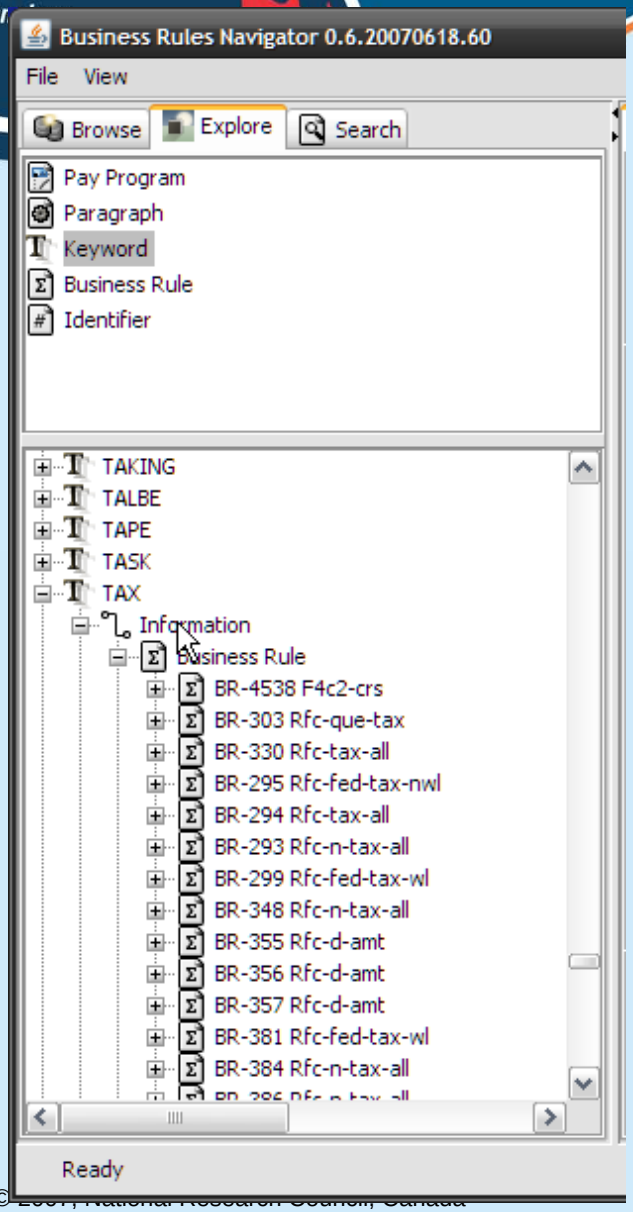
Basic navigation



- **Hierarchical navigation**
  - Programs
  - Paragraphs
  - Business Rules

NRC·CNRC
Institute for Information Techn

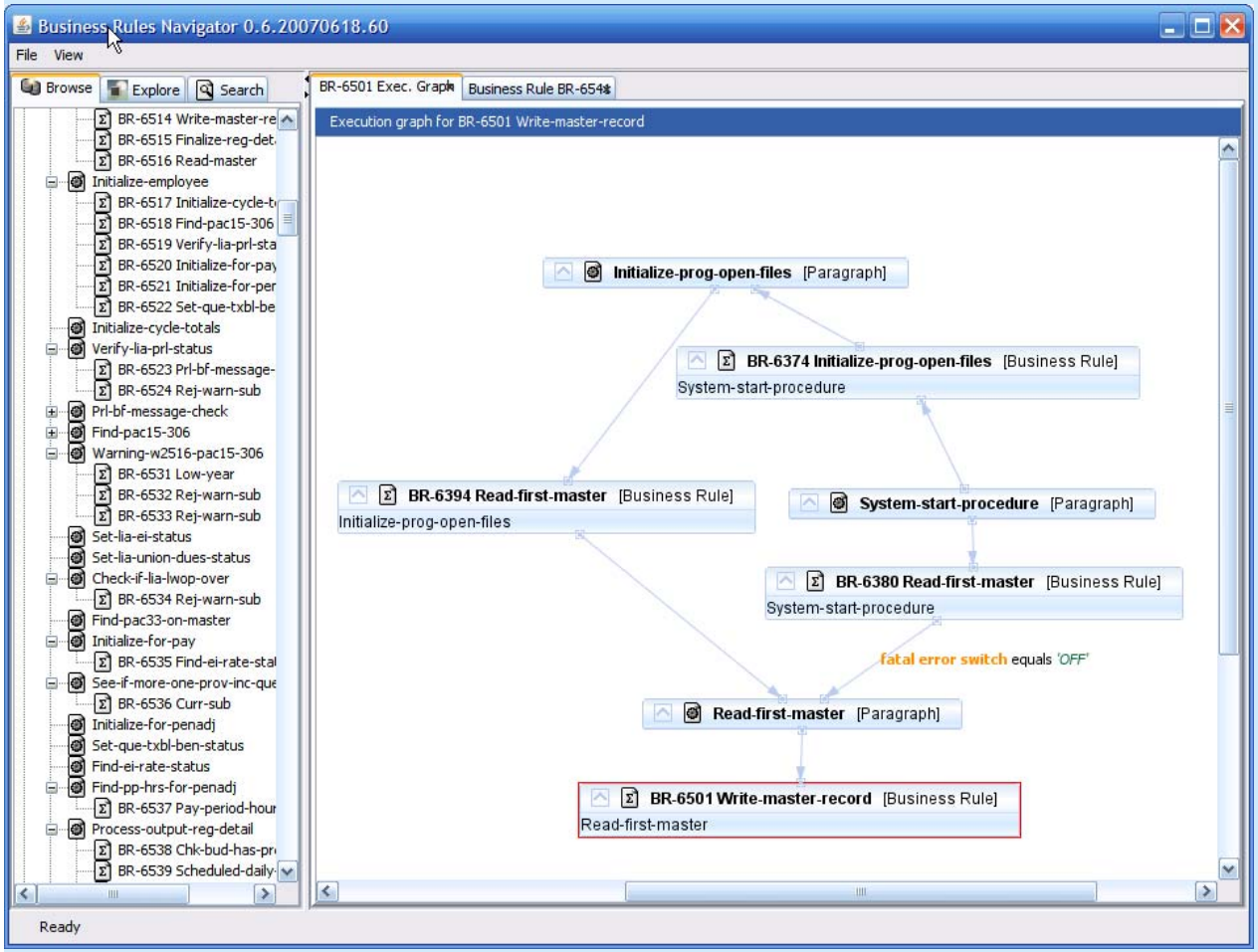
System wide navigation



- **Keyword indexing**
  - Identifiers, comments and COBOL locations are converted to keywords and indexed

NRC·CNRC
Institute for Information Technology

Execution path and dependency visualization



- **Show all paths leading to the execution of one Business Rule**

NRC·CNRC
Institute for Information Technology

Outline

- Context and objectives
- Extraction process
- Visualization and navigation
- **Conclusions**

NRC·CNRC
Institute for Information Technology

Conclusions

- **Business Rules: major element in legacy software modernization**
  - For system maintainers and business analysts
- **Possible to extract business rules from legacy source code**
- **Novelty**
  - output is targeted at business analysts
  - the business rules translated into non-technical terms.
- **Positive early feedback from business analysts about the value and understandability of the information extracted**

NRC·CNRC
Institute for Information Technology

Future

- **Simplified information about branching and loops**
- **Integrate the documentation into the system**
  - Link all elements of the database with existing documentation
- **Data flow analysis**
  - From a generated report of the legacy system, locate all related business rules and calculations
  - State analysis through the execution flow for connecting variables to documented data sources