

МГТУ МИРЭА

Выпускная квалификационная работа

Разработка программного обеспечения
для извлечения, обработки и верификации
полнотекстового содержания и
метаданных библиотечных электронных
единиц хранения.

Выполнил Воронин А.А. КБ-86-09

Научный руководитель к.т.н., доцент Парфенов Д.В.

Консультант к.т.н., доцент Федоров В.Б.

Постановка задачи

Рассмотрена задача извлечения полнотекстового содержания и метаданных из книг электронной библиотеки в форматах PDF, PostScript (PS), DJVU, CHM и HTML; разработка системы автоматизированной пакетной обработки электронных единиц хранения библиотеки в этих форматах с целью выделения следующих элементов информации:

1. полный текст документа без управляющих и форматных символов;
2. название документа;
3. список авторов;
4. количество страниц;
5. год издания;
6. язык документа;

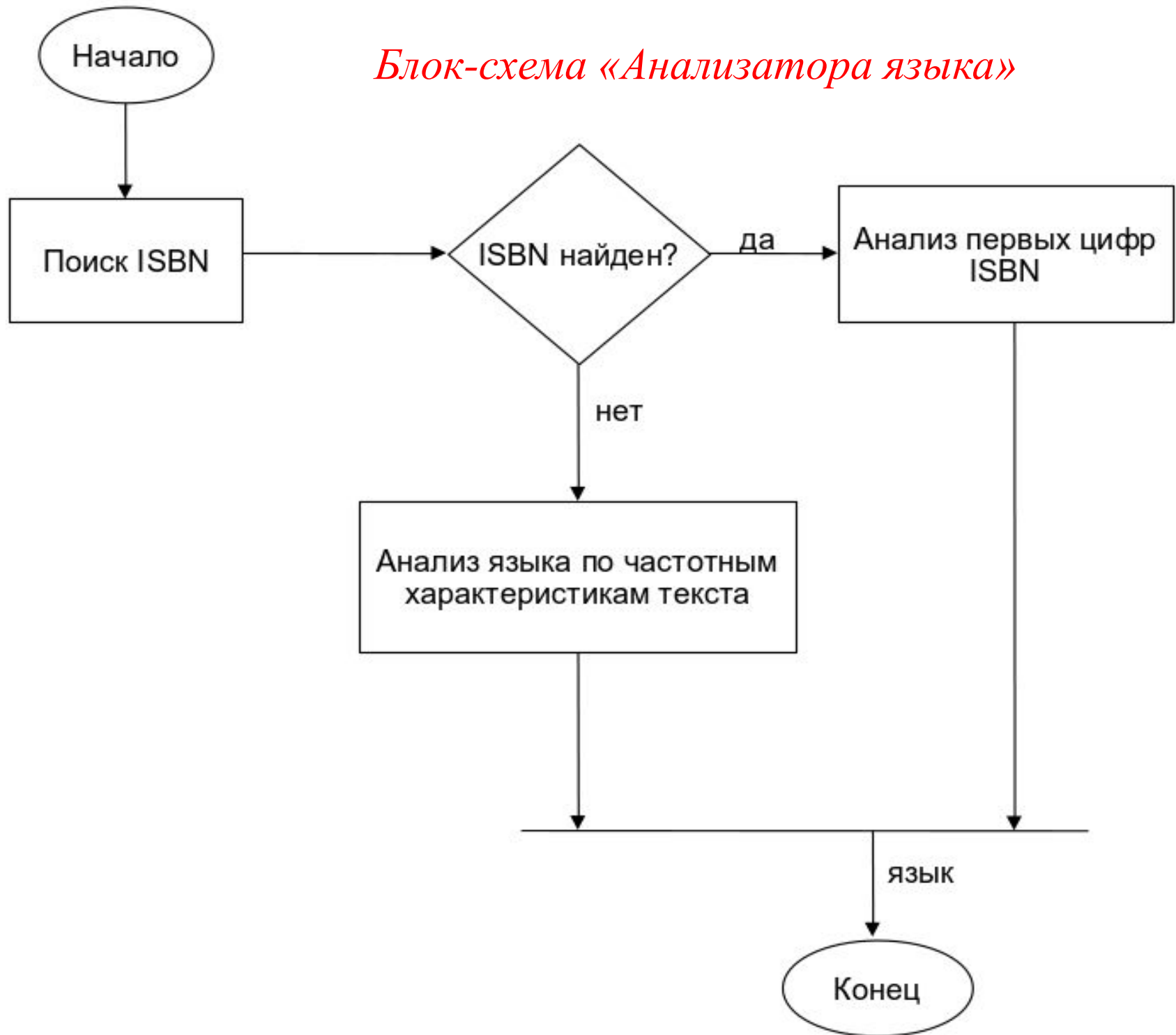
Title: Jesus and His Death : Historiography, the Historical Jesus, and Atonement Theory
Location: file:///home/uno/diplom/engB/a4
Subject: *None*
Author: McKnight, Scot.
Keywords: *None*
Producer: QuarkXPress(tm) 6.5
Creator: QuarkXPress(tm) 6.5
Created: Ср. 29 июня 2005 14:28:21
Modified: Сб. 07 апр. 2007 22:22:00
Format: PDF-1.4
Number of Pages: 461
Optimised: Yes
Security: No
Paper Size: 156 × 235 mm

Пример метаданных электронной книги

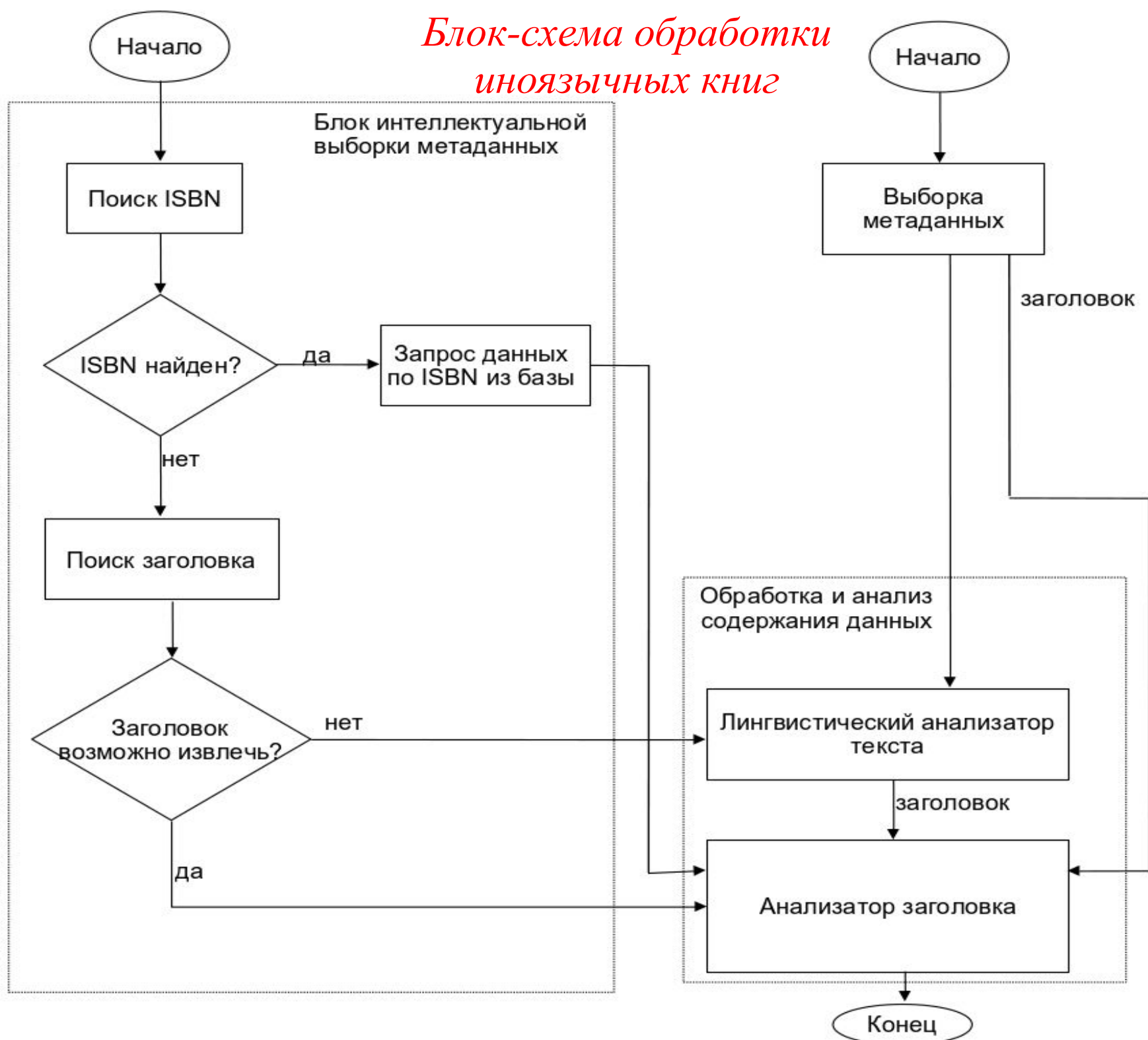
Блок-схема системы извлечения метаданных

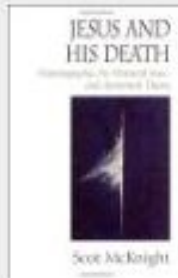


Блок-схема «Анализатора языка»



Блок-схема обработки иноязычных книг





[Jesus and His Death](#)

Scot McKnight

Best Price \$41.74
or Buy New \$41.74



[Privacy Information](#)

Jesus And His Death: Historiography, the Historical Jesus, And Atonement Theory

McKnight, Scot

Publisher: Baylor University Press

ISBN10: 1932792295 ISBN13: 9781932792294 Edition: Hardcover; 2005-09-30

Book Details

Physical Description: 451 pages

Edition Info: Hardcover; 2005-09-30

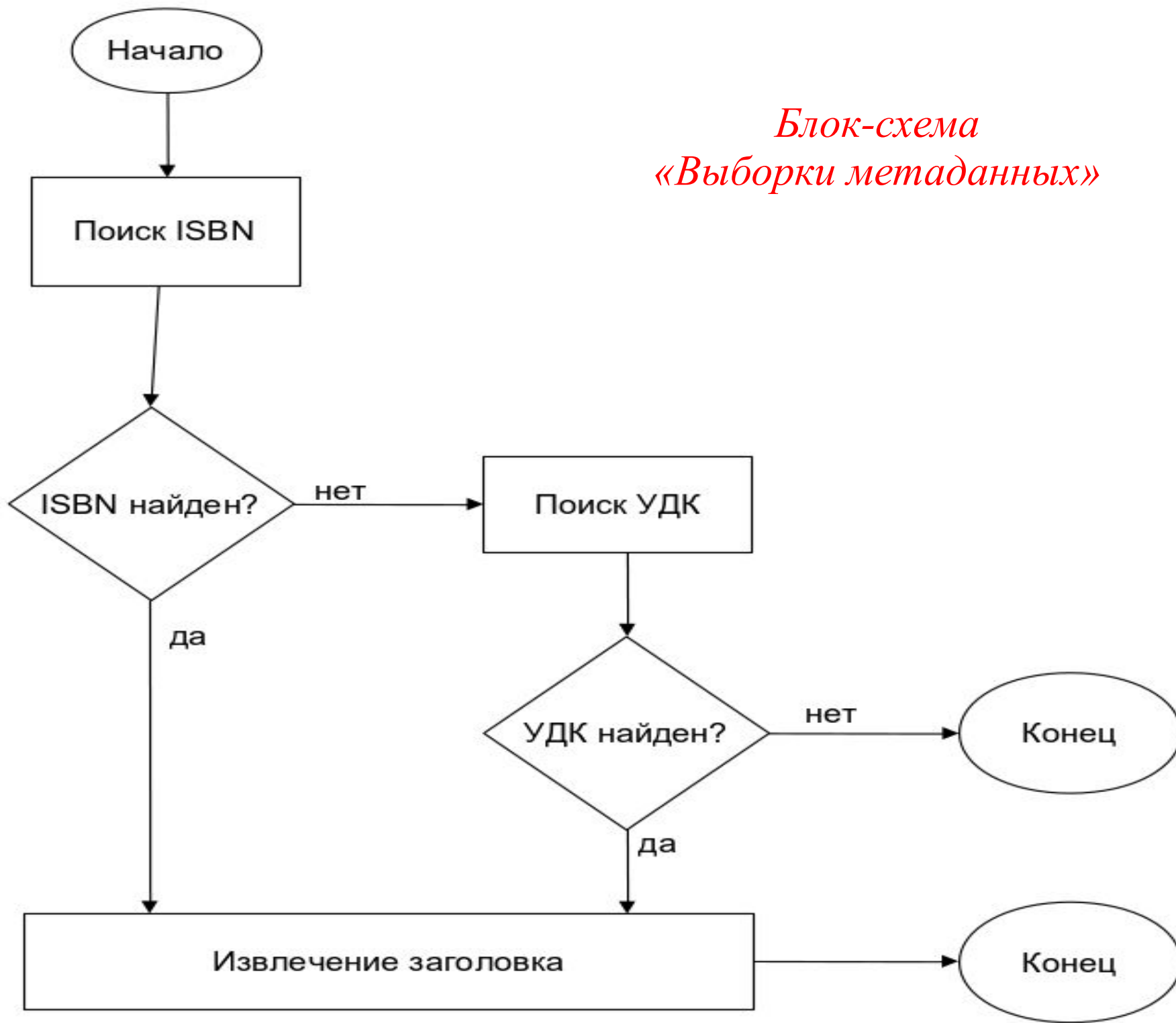
Buying This Book

We query many merchants so that you can instantly **compare prices and availability**. You can even check historic prices and subscribe for notifications. For a manual check, clicking on a link will open a new window with a search for this book on the merchant's site of your choice.

- Amazon.com by **title** or by **ISBN** - widest selection, reliable service, good prices
- Abebooks.com by **title** or by **ISBN** - good source for rare and out-of-print books
- Textbooks.com by **title** or by **ISBN** - large inventory of new & used textbooks
- Alibris.com by **title** or by **ISBN** - their slogan is "books you thought you'd never find"
- eCampus.com by **title** or by **ISBN** - they specialize in textbooks, used and new
- BN.com by **title** or by **ISBN** - Barnes & Noble has a wide selection of new and used books

Пример отклика сайта isbndb.com на запрос

*Блок-схема
«Выборки метаданных»*



УДК 53(075.8)
ББК 22.3я73
Т761

*Элемент текста с метаданными
в русскоязычной книге*

Рецензент —
профессор кафедры физики им. А. М. Фабриканта
Московского энергетического института (технического университета) *В.А. Касьянов*

Трофимова Т. И.

Т761 Курс физики: учеб. пособие для вузов / Таисия Ивановна Трофимова. — 11-е изд., стер. — М.: Издательский центр «Академия», 2006. — 560 с.
ISBN 5-7695-2629-7

uno@uno-x: ~/diplom/engB

File Edit View Search Terminal Help

uno@uno-x:~/diplom/engB\$ sh allComm

--2013-06-26 14:40:02-- http://isbndb.com/search/all?query=1932792295

Resolving isbndb.com (isbndb.com)... 64.27.101.156

Connecting to isbndb.com (isbndb.com)|64.27.101.156|:80... connected.

HTTP request sent, awaiting response... 302 Found

Location: /book/jesus_and_his_death [following]

--2013-06-26 14:40:02-- http://isbndb.com/book/jesus_and_his_death

Reusing existing connection to isbndb.com:80.

HTTP request sent, awaiting response... 200 OK

Length: 9264 (9,0K) [text/html]

Saving to: `all?query=1932792295.3`

100%[=====

2013-06-26 14:40:03 (53,2 KB/s) - `all?query=1932792295.3' saved [9264/9264]

Jesus And His Death: Historiography, the Historical Jesus, And Atonement Theory

mcknight scot

Baylor University Press

451 pages

Hardcover; 2005-09-30

uno@uno-x:~/diplom/engB\$

*Пример текстового
вывода программы*

Спасибо за внимание!