



Development of Personal Income Tax Management System Based on Data Mining Technology

Yiran Tao, Yixi Zhang, and Xiaoqian Wang^(✉)

Chongqing College of Architecture and Technology, Chongqing 401331, China
102314404@qq.com

Abstract. In the current environment of rapid development and application of network information technology, based on data mining technology, combined with Web application development technology under the J2EE specification, a personal income tax management system is constructed, which is the issue of personal income tax collection and management under the new tax system. Propose comprehensive solutions for digitalization and informatization. Based on the actual needs of both taxpayers and tax authorities, the system will combine the collection and administration methods and process of individual income tax, with functional modules such as declaration management, personal information management, analysis and evaluation, etc., to realize the online innovation of individual income tax payment services, improve the It improves the efficiency of personal income tax collection and management, reduces the cost of collection and management, and provides scientific data support for the formulation of tax system decisions, promotes the further improvement of tax collection, management, and audit systems, gives full play to the due functions of personal income tax, and promotes the realization of taxation in the whole society. Equity, and promote stable economic and social development.

Keywords: personal income tax · data mining · tax collection and management · J2EE · informatization

1 Introduction

Over the past 40 years, China has adhered to the fundamental policy of economic construction with reform and opening up, and has made remarkable achievements. As of 2021, the gross domestic product (GDP) is 114 trillion yuan, which has surpassed the European Union to become the second largest economy of the century. Domestically, absolute poverty in rural areas has been completely eliminated, and the overall income level of the people has been raised to the greatest extent possible. However, it should also be noted that while Chinese economy is growing rapidly, the diversification of income channels, the diversification of income patterns, the readjustment of the pattern of interests during the reform process, and the widening gap between the rich and the poor have become a focus of our current society question [7]. Faced with this situation, on the

one hand, the state should continue to develop the economy and expand employment to increase people's income, and on the other hand, it should improve the redistribution mechanism by means of taxation, promote social fairness and justice, and promote the goal of comprehensive development and common prosperity. accomplish. In my country's current tax system structure, personal income tax, as a typical direct tax, can avoid the impact of tax burden transfer compared with indirect tax. By directly affecting the disposable income of individuals, the goal of narrowing the gap between the rich and the poor is achieved.

The personal income tax has been collected and paid since 1980. In the course of more than 30 years of development, it has undergone several adjustments in multiple stages to adapt to the changes in the income gap of the Chinese people. On January 1, 2019, the latest Individual Income Tax Law was officially implemented. The main contents of this reform include transforming the previous classified tax system into a comprehensive tax system, including four items of income from wages and salaries, labor remuneration, author's remuneration and royalties. The scope of taxation, and the corresponding implementation of the corresponding collection and management model. Raise the threshold and increase 6 special additional deductions to further reduce the tax burden of the people. Optimizing and adjusting the tax rate structure of personal income tax will make Chinese personal tax system more scientific and rational, and give full play to the regulatory role of personal income tax [8]. The new personal income tax system is a tax system reform with the largest adjustment scale, the widest range of people involved, and the strongest tax reduction since its establishment. a whole new challenge. First of all, the management object has changed from the previous withholding unit to the withholding unit and the natural person taxpayer in parallel, and the natural person taxpayer group has a large base, strong liquidity, and weak tax awareness, which makes it difficult for tax authorities to control and manage individual income tax sources. Secondly, the coverage of tax-related information of natural persons has increased, and there are multiple income channels, which are discrete and concealed, which increases the difficulty of obtaining tax-related information. For complex special additional deduction information, there is a lack of means to quickly verify the information. Finally, the declaration information of individual income tax will be assessed, audited and inspected by the tax authorities to form a collection and management model of prior reminder, in-process verification and post-assessment. In view of this, this paper believes that, with the help of the application advantages of network information technology, with the Web application development technology under the J2EE specification as the core, combined with the current collection and management method and collection process under the new personal income tax system, the construction of the personal income tax management system is completed. On the one hand, the system can meet the needs of natural person taxpayers to obtain convenient and efficient online tax payment services; The flow and interaction speed in the verification stage can facilitate the transmission and call between different departments only through the data interface. Finally, the system also introduces data mining technology into after-tax assessment, and uses scientific data analysis results to enhance the management and control of tax collection and management information data, improve the professional degree of personal income tax source management, and provide necessary data for the formulation of tax policies. support.

2 Introduction of Key Technologies

We strongly encourage authors to use this document for the preparation of the camera-ready. Please follow.

Please remember that all the papers must be in English and without orthographic errors.

2.1 Data Mining Technology

Data Mining (DM) refers to the analysis of the massive data information stored in the database, to find the relationship and value between the data information, in order to discover the previously unknown, implicit and useful knowledge and laws for human beings. complex process. The birth of data mining technology meets people's needs for analysis and processing of massive information data in the current information explosion era. Technology [10]. Compared with traditional data analysis, data mining technology has the ability to discover knowledge, that is, through current data information, through "thinking" to obtain unknown knowledge, and it is extremely efficient and convenient. In addition, data mining technology can deal with more complex data objects, including structured data, semi-structured data and heterogeneous data, including multiple data types or data sources.

2.1.1 Data Mining Process

The main process is shown in Fig. 1, in which databases, data warehouses, the Internet and other data sources represent the collection of raw data. The collected raw data contains numerous data objects, with different data types, data formats, and obvious irrelevant data or noise data, which requires data preprocessing to adjust and standardize the data for the results of subsequent data mining. Provide quality assurance. Preprocessing includes three basic steps: data cleaning, data integration, data selection and transformation. Among them, data cleaning will remove the "dirty data" that affects the results of data mining; data integration can combine relevant data from multiple data sources and store them in a unified manner; selection and transformation transforms data into a unified form suitable for mining, improving follow-up operational efficiency. The preprocessed data is stored in the database, waiting for subsequent data mining operations to be called. After users enable data mining technology, they will select different mining methods according to their own requirements, establish an effective data mining analysis model based on data attributes and data warehouses, complete real data mining operations, and visualize the results of data mining. Show it, improve human-computer interaction, and achieve rapid grasp of knowledge and information.

2.1.2 Data Mining Methods

As the core of data mining technology, the choice of data mining methods is related to the success or failure of building data mining analysis models, and is the key to the

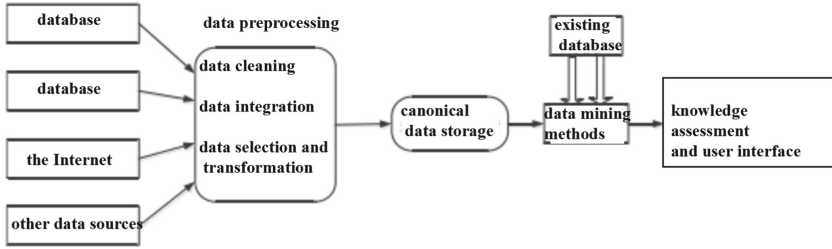


Fig. 1. Data mining process

results of data mining. At present, the most widely used and most mature data mining methods are correlation mining, classification mining and clustering mining. Combined with the application requirements of tax authorities for the assessment and auditing of taxpayers’ tax-related information under the personal income tax management system in this article, the method of classification mining will be used, and the decision tree classification algorithm will be used to preprocess tax data and tax-related information, and establish decision-making. Tree tax assessment, audit case selection assessment models, generate decision trees, and provide decision support for actual tax collection and management.

2.2 Decision Tree Algorithm and C4.5 Algorithm

The decision tree algorithm belongs to the classification mining algorithm, which is an inductive learning algorithm that can classify instances. The algorithm is applied to unordered and irregular instances. It is constructed in a top-down recursive way. An internal node represents a feature or attribute, and a leaf node represents a class. Each path from the top to the leaf node is a The classification rules are integrated to form a tree-structured expression prediction analysis model, as shown in Fig. 2. The decision tree algorithm adopts the recursive strategy of “divide and conquer”. The purpose is to further divide the complex problem into simpler and smaller sub-problems, and combine the solutions of the sub-problems to generate the solution of the complex problem and realize the processing of the complex problem [6]. Therefore, the decision tree classification algorithm has obvious advantages such as fast speed, high accuracy and strong continuity, and is more suitable for the application requirements of this system.

The key of the decision tree algorithm in inductive learning is to select the optimal division attribute, that is, the division and control of the internal nodes in the above. After the internal nodes (features) of each layer are selected, the decision tree is naturally built. There are many division standards for internal nodes, the common ones are ID3 algorithm and C4.5 algorithm. The difference between the two algorithms lies in the use of information gain or information gain rate as the division criterion (branch criterion). The C4.5 algorithm uses the information gain rate to evaluate the internal nodes that may be divided, as shown in Eq. 1, where D is a sample set, a is a discrete attribute, and Gain(D, a) is the attribute a. The information obtained by dividing the sample D gain, IV(a) is the intrinsic value of attribute a. The greater the information gain rate, the more useful information the branch contains, the more important the attributes or features

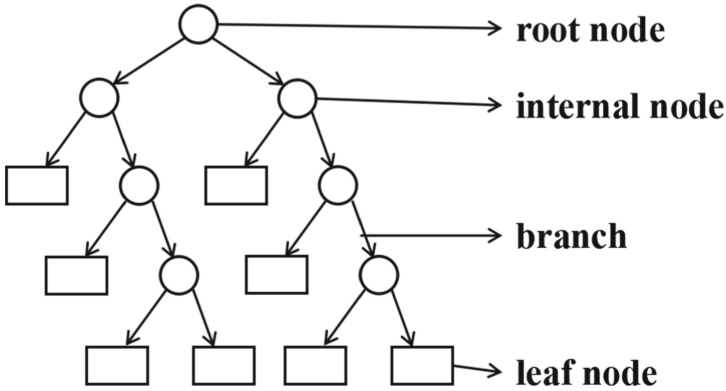


Fig. 2. Decision tree algorithm structure diagram

represented by this internal node, and the more accurate the overall decision tree model [4].

$$\text{Gain_ratio}(D, a) = \frac{\text{Gain}(D, a)}{\text{IV}(a)} \tag{1}$$

Compared with the ID3 algorithm, the C4.5 algorithm discretizes the continuous features, which can increase the flexibility, usability and efficiency of the decision tree algorithm, and is more suitable for the needs of the personal income tax system in this paper.

2.3 J2EE

J2EE is the acronym for Java 2 Platform Enterprise Edition (Java 2 Platform Enterprise Edition). Its core is a set of comprehensive technical specifications, including Java Servlets, Jsp pages, Java Beans, etc. Its purpose is to provide platform-independent, portable, support for concurrent access and security, completely Java-based development server-side. The standard of middleware can make good compatibility between different platforms of J2EE architecture, which ensures that the information products used by different enterprises can be unimpeded with each other [3]. In the actual development and application process, J2EE adopts a multi-layer distributed application model to provide an independent layer for each functional service in the system to increase the scalability and reusability of the system. As shown in Fig. 3, it is the J2EE distributed application model structure, in which the client tier (Client tier) runs on the remote client to complete the communication with the J2EE server by the HTTP protocol, and the application components are generally HTML pages. The WEB tier runs on the J2EE server. The components mainly include JSP pages or Servlets. Both JSP and Servlets can obtain requests from users in time, and forward the requests to Enterprise Beans on the business tier for processing. In the business layer (EJB tier), the corresponding business processing is completed by the Enterprise bean in the EJB container. The final enterprise information system layer (ESI tier) runs on the database server side, emphasizing the storage and invocation of various data information resources in the system. The

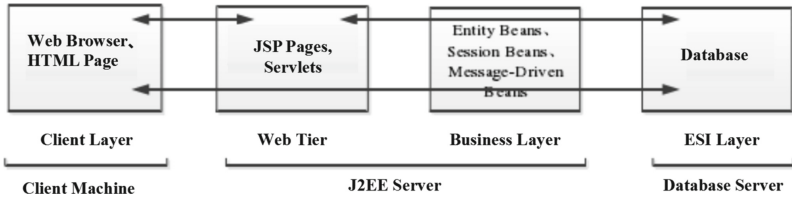


Fig. 3. J2EE distributed application model structure

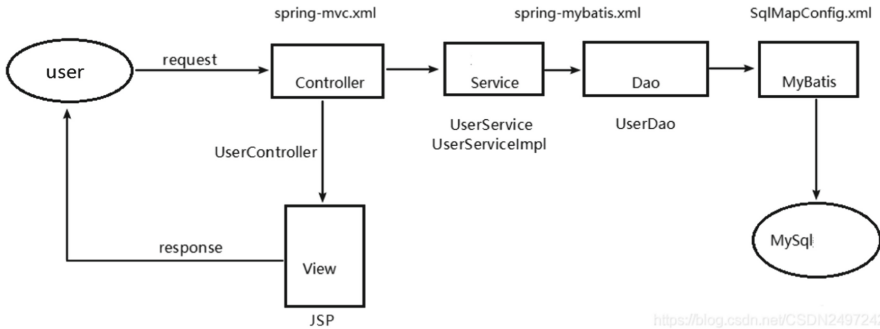


Fig. 4. SSM framework running process

application of J2EE can provide the necessary functional support for the development of this system, and it also has significant advantages such as strong scalability, stable availability, efficient development, and easy maintenance.

2.4 SSM Framework

The SSM framework is an abbreviation for the three basic frameworks of Spring, Spring-MVC, and MyBatis, and is also an integrated framework suitable for web application development. The SSM framework is composed of two open source frameworks, Spring and MyBatis, while SpringMVC belongs to the Spring framework and is part of its framework. The SSM architecture is subordinate to the J2EE development specification. Compared with the same type of SSH framework, it is more lightweight and flexible. It can simplify the process links of Web applications and realize the agile development of the system with its good reusability. As shown in Fig. 4, the flow chart is run for the SSM framework, in which Spring MVC is the core of the SSM framework and is constructed by the underlying technologies such as Servlet in J2EE. It takes the Dispatcher Servlet as the core and coordinates components such as Handler Mapping, View Resolver, and Handler Adapter to complete the response of HTTP requests [1].

2.5 Development Environment

Complete the configuration and deployment of the development environment according to the system development requirements and the use requirements of the above-mentioned key technologies. The overall development of the system is based on the Windows 10.0 operating system, the basic development environment is Java, the JDK version requires 1.8 and above, the Java development environment is MyEclipse2017, the web server is Tomcat 8.5, and the database server is MySQL 5.7. In the process of building the overall development environment, use Maven to manage the project structure, split the entire project into multiple engineering modules, and complete the integration of Spring - Spring MVC - MyBatis based on the SSM architecture. Create a new Web project under MyEclipse2017, and then import the Jar packages required for development, including

```

<context-param>
  <param-name>contextConfigLocation</param-name>
  <param-value>classpath:spring-mybatis.xml</param-value>
</context-param>
<filter>
  <filter-name>encoding</filter-name>
  <filter-class>org.springframework.web.filter.CharacterEncodingFilter</filter-class>
  <init-param>
    <param-name>encoding</param-name>
    <param-value>UTF-8</param-value>
  </init-param>
</filter>
<filter-mapping>
  <filter-name>encoding</filter-name>
  <url-pattern>*.action</url-pattern>
</filter-mapping>
<listener>
<listener-class>org.springframework.web.util.IntrospectorCleanupListener</listener-class>
</listener>
<listener>
<listener-class>org.springframework.web.context.ContextLoaderListener</listener-class>
</listener>
<servlet>
  <servlet-name>SpringMVC</servlet-name>
  <servlet-class>org.springframework.web.servlet.DispatcherServlet</servlet-class>
  <init-param>
    <param-name>contextConfigLocation</param-name>
    <param-value>classpath:spring-mvc.xml</param-value>
  </init-param>
  <load-on-startup>1</load-on-startup>
  <async-supported>true</async-supported>
</servlet>
<servlet-mapping>
  <servlet-name>SpringMVC</servlet-name>
  <url-pattern>/</url-pattern>
</servlet-mapping>
</web-app>

```

Fig. 5. Key code under the web.xml configuration file

the database connection package `mysql-connector-java`, `mybatis` core package, `spring` core package, `mybatis-spring` package, etc. After completing the relevant configuration, modify the configuration file under `web.xml`. The key code is shown in Fig. 5, where `classpath:spring-mybatis.xml` is the listener for the system to load the Spring file to obtain the user's request. `Org.springframework.web.filter.CharacterEncodingFilter` is the declaration and configuration of the encoding filter. Declare `<servlet>` as the Spring MVC front-end core controller to handle user requests. Through the introduction of the above key technical theory, the overall environment of system development, the configuration of related software and tools are determined, and the technical feasibility of the overall project of the personal income tax management system is also clarified.

3 Demand Analysis

We strongly encourage authors to use this document for the preparation of the camera-ready. Please follow.

Please remember that all the papers must be in English and without orthographic errors.

3.1 System Requirements Analysis

Based on the Java Web application development ideas under the J2EE specification, the system developers take the actual collection and management process as the design blueprint, and combine the current taxpayers and tax authorities' needs for personal income tax collection and management services to design and develop a personal income tax management system. The purpose is mainly It is to establish a complete and effective information system that integrates individual income tax declaration, management, evaluation, inquiry and service, emphasizing the system's ability to generate taxpayers' individual tax declaration information data, secure data transmission, data auditing and mining analysis to comprehensive sharing. Various functional realizations and technical support such as application, etc., enhance the ability of tax authorities to monitor and utilize complete and accurate personal income data [5].

The system can support the two user roles of taxpayer and taxpayer to complete the login and use of the system through account registration. The system will set up the taxpayer subsystem and taxpayer subsystem according to the actual needs and application permissions of users with different roles. Under the taxpayer subsystem, the main functions include the declaration and management of personal income tax, and the management of basic information. On the one hand, it is convenient for taxpayers to complete personal income tax declaration operations through online services, and on the other hand, it also provides taxpayers with a comprehensive storage space for personal tax information, which is convenient for subsequent inquiries and calls, and provides a data basis for subsequent evaluation management. The main function under the taxation personnel subsystem is to analyze and evaluate the taxpayer's personal income tax declaration information. The system has embedded data mining function, which can support taxation personnel to complete the review, analysis, confirmation and evaluation through simple operations. In order to timely detect, correct and deal with abnormal situations in tax declaration behaviors, and urge taxpayers to pay taxes in accordance with the law.

3.2 Overall Design

The personal income tax management system will adopt B/S structure, under the J2EE technical specification, using Java Web application development technology as the foundation, combined with the SSM framework to complete the overall design and development of the system. This system adopts B/S three-layer structure, and divides the system into presentation layer, business layer and data layer. The MVC design can be implemented for the presentation layer by using the Spring MVC framework, and the transaction processing mechanism supported by the Spring framework is used under the business layer. Finally, the Mybatis framework is used in the data layer to establish the mapping between entity classes and database tables, which provides the basis for the realization of the functions of the business layer [2]. When taxpayers or tax officials log in to the system with any browser connected to the Internet and submit a request to the server, Servlet or JSP generates web pages to complete the presentation layer logic, they accept and verify user input, and then call Javabeen to perform For business logic processing, access the database through JDBC, and then return the data and business logic processing results to the server, and return them to the client in the form of HTML pages (Fig. 6).

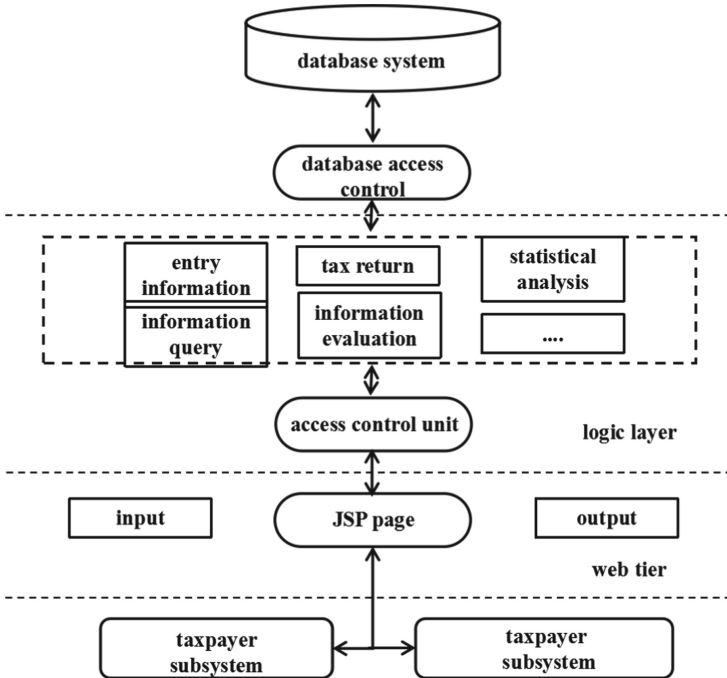


Fig. 6. Overall system architecture diagram

4 Detailed Function Implementation

4.1 Taxpayer Subsystem

In the personal income tax management system, the taxpayer is the initiator of the use of the entire system, which is related to the capture and integration of the overall data and information resources of the system. When a taxpayer user logs in to the system for the first time, he or she needs to complete the account application and registration according to the operation guide. Once approved, you can complete the login with a unique username and password. The main functions of the taxpayer subsystem include declaration management and basic information management.

Under the reporting management function, the system supports users to complete operations such as information entry, filling in reports, reporting and feedback. Under the information entry function, taxpayers can add their personal details to facilitate the establishment of a one-household file for personal income taxation. Under the function of filling in the report, the user can fill in the personal tax information according to the personal income tax self-tax return (Form A) or the personal income tax self-tax return (Form B) provided by the system, including the total income, total expenses, and total tax-exempt income., total special deductions, total special additional deductions, and total other deductions, etc. After the user completes the filling, the system will automatically generate the declaration and reconciliation data and complete the submission. After the submission, the system will automatically jump to the receipt interface, and the user can click to download the receipt for this application.

Under basic information management, users can inquire about personal basic information and tax information, and can also analyze relevant tax data by type and time, and support data export or download and export.

4.2 Taxpayer Subsystem

Under this subsystem, tax personnel will analyze and compare individual tax declaration information, tax-related information and key indicators, and complete the authenticity judgment of the declaration situation through data mining technology, which is the current individual tax system under the new tax system. Tax collection and management provide the necessary basis. The overall process is carried out in the order of five steps: identifying objects, evaluating and analyzing, drawing up lists, evaluating and processing, and managing suggestions. The core of the individual income tax analysis and evaluation module is the establishment of the tax evaluation index system, and the tax evaluation objects are screened accordingly [9]. First, the tax personnel select the sample space according to the basic information of the taxpayer user and the tax declaration information in the system, and determine the evaluation object. Evaluation analysis is to use the C4.5 algorithm to construct a decision tree analysis model, and to evaluate and analyze all the selected evaluation objects. The result of evaluation analysis indicates the classification of all evaluation objects, one is qualified classification and the other is suspicious classification. Under the suspicious classification, a corresponding object list is drawn up, and all the evaluation objects on the list are evaluated and analyzed again after manual review or necessary inquiry and verification, and the final evaluation

and processing results are obtained. Tax personnel will conduct follow-up tax audits and inspections based on the results of the assessment and processing.

5 Conclusions

The personal income tax management system based on data mining technology can take advantage of the application advantages of current network information technology and the application expertise of data mining technology in data analysis and processing, and propose a comprehensive and applicable method for the collection and management of personal income tax under the current new tax system. solution. The system can not only satisfy the taxpayers to complete the declaration and submission of personal tax through the online form, but also facilitate the tax personnel to improve the office efficiency of the tax collection and management. The application of the system not only improves the quality of tax services, but also enriches the forms of tax services. While strengthening the scientific and comprehensive supervision and management of taxpayers' individual tax declaration behavior, it also effectively promotes the network and information in the field of tax collection and management. development process.

References

1. Chen Heng, Lou Oujun. (2021). Spring MVC Development Technical Guide. Computer Education.
2. Chen Zhao. (2021). Design and Implementation of SSM Framework in Web Application Development. Computer Knowledge and Technology.
3. Ke Cai. (2020). Research on Web Application Based on J2EE and MVC Pattern. Computer Engineering & Software.
4. Han Cunge, Ye Qiusun. (2019). Research and improvement of C4.5 algorithm in decision tree classification algorithm. Computer System Application.
5. He Xue. (2020). Research on the information collection and management of personal income tax for natural persons. Tianjin University of Finance and Economics.
6. Li Yingchun. (2015). Research on decision tree classification algorithm in data mining. Hunan Normal University.
7. Wang Jing, Deng Xiaolan. (2021). Evaluation of income distribution effect of my country's personal income tax reform. Journal of Xi'an University of Finance and Economics.
8. Wang Jun. (2021). The current difficulties and countermeasures of personal income tax collection and management. Journal of Hunan Taxation College.
9. Yang Hui. (2018). Design and implementation of personal income tax management system in Qinzhou. University of Electronic Science and Technology of China.
10. Zhang Bo. (2020). Data mining technology and application in the era of big data. Digital Technology and Application.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

