

ОБ ЭЛЕКТРОННОЙ КОЛЛЕКЦИИ ФОЛЬКЛОРНЫХ ПЕСЕН С ТЕОРЕТИКО-ГРАФОВОЙ ФОРМАЛИЗАЦИЕЙ ТЕКСТОВ

А.Г.Варфоломеев, Н.Д.Москин (Петрозаводский государственный университет, avarf@mainpgu.karelia.ru)

В современном Интернете можно выделить три типа информационных систем для историков: справочно-библиографические, статистические и полнотекстовые. Статистические и полнотекстовые информационные системы в большинстве своем содержат оцифрованные первичные источники и, таким образом, исполняют роль небольших специализированных архивов. Такие системы все чаще создаются не для решения какой-то конкретной научной проблемы, а прежде всего с целью адекватного представления архивных данных в электронной форме, что определяет их источник-ориентированный характер, а также постановку, наряду с научными, и культурно-просветительских задач.

Такой подход оказывается оправданным прежде всего с научной точки зрения, так как значительно увеличивает возможности повторного использования данных другими исследователями. Однако, создавая источник-ориентированную базу данных, не стоит забывать и о следующем естественном этапе работы – формализации и генерализации информации.

Авторами доклада ведется работа над проектом по созданию информационной системы “Бесёдные песни Заонежья конца XIX – начала XX вв.”. За основу взята уже существующая полнотекстовая база данных фольклорных песен, собранная старшим научным сотрудником музея “Кижы” Р.Б.Калашниковой[1]. Для повышения исследовательского потенциала информационной системы необходимо решить задачу формализации содержания песен.

Для формализации текстов учёные-фольклористы часто применяют типологические формулы. При этом формализация одновременно решает и задачу классификации объектов: в качестве их признаков как раз и выступает принадлежность объектов тому или иному классу. Однако такого рода “ручная” классификация не позволяет выявить скрытую от исследовательского взгляда информацию. Хотелось бы описать фольклорную песню такой математической структурой, которая непосредственно не задавала бы классификацию объектов, но могла бы обрабатываться с помощью алгоритмов автоматической классификации.

На наш взгляд, наиболее адекватной математической структурой, описывающей фольклорную песню, служит множество ориентированных графов, связанных между собой. В каждом графе узлами являются объекты текста, а дугами – отношения между объектами. В роли объектов текста выступают главным образом существительные и прилагательные (эпитеты), отношения выражаются, в частности, глаголами. Графы отражают сюжетную композицию песни, каждому графу соответствует некая сюжетная линия (“мотив”).

В докладе представлен прототип информационной системы, которая будет содержать как полные тексты песен, так и их формальные представления в виде графов, что позволит обоснованно проводить сравнения песен, выделять инварианты, решать задачи классификации фольклорного материала, и, кроме того, использовать уникальную историческую информацию в справочных, учебных, культурно-просветительских целях.

Литература

1. Калашникова Р.Б. Бесёды и бесёдные песни Заонежья второй половины XIX века. Петрозаводск, 1999.

THE ELECTRONIC COLLECTION OF FOLKLORE SONGS WITH TEXTS FORMALIZATION BY GRAPHS

Aleksey G. Varfolomeyev, Nikolay D. Moskin
Petrozavodsk State University
Petrozavodsk, Russia

The authors elaborate the project of information system dedicated to folklore songs of North Russia. For increase of research potential of information system it is necessary to decide a task of formalization of the song content. In our opinion the most adequate mathematical structure describing a folklore song is the set of oriented graphs connected among themselves. In each graph the vertexes are the objects of the text, the edges are relations between objects. The each plotline of song corresponds to certain graph. Our information system will contain both complete texts of songs and their formal representations by graphs that will allow to carry out comparisons of songs, to find invariants, to decide tasks of classification.