

УДК 519.711.2

## ОПЫТ БИКЛАСТЕРИЗАЦИИ ДАННЫХ О СОРТАХ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР

© 2020 В.А. Семенова<sup>1,2</sup>

<sup>1</sup> Институт управления сложными системами Российской академии наук – обособленное подразделение Самарского федерального исследовательского центра РАН, г. Самара

<sup>2</sup> Самарский государственный технический университет

Статья поступила в редакцию 30.12.2019

Статья посвящена анализу объектно-признаковых данных в слабо структурированных областях. Конкретно, рассматривается демонстрационный пример бикластеризации подобных данных в области селекции сельскохозяйственных культур. Показывается, какие при этом открываются возможности для систематизации эмпирических данных. Рассматриваемый метод бикластеризации данных – классический анализ формальных понятий – представляет собой прикладную ветвь теории решеток; приведена краткая справка о его математических основах. Исходный материал для демонстрационного примера был заимствован из каталога инновационных разработок Самарского НИИСХ им. Н.М. Тулайкова. В качестве объектов анализа избраны сорта мягкой яровой пшеницы, в качестве признаков – несколько описывающих их свойств. Ряд этих свойств подвергнут шкалированию в соответствии с ГОСТ 9353-2016 для пшеницы. На основе отобранного и скорректированного материала сформирована исходная таблица «объекты-свойства», или формальный контекст для анализа формальных понятий. Для обработки этих данных использован свободно распространяемый программный пакет ConExp, реализующий методический комплекс анализа формальных понятий. Продемонстрированы основные результаты, которые дают эти методы для структуризации эмпирических данных: решетка формальных понятий, импликации и ассоциативные правила на множестве признаков. Кроме того, продемонстрирован результат работы программного комплекса OntoWorker, разрабатываемого в ИПУСС РАН – СамНЦ РАН, заключающийся в трансформации редукированной решетки формальных понятий в специальную таксономию формальных понятий (классов), более удобную для восприятия и интерпретации пользователем. Отмечено, что расширенные методы анализа формальных понятий могут учитывать неполноту исследуемых данных, обусловливаемую множественностью серий измерительных экспериментов с разной степенью достоверности результатов и наличием конкурирующих измерительных процедур с разной степенью доверия к их результатам. Это отражает реалии накопления эмпирической информации и может быть использовано непосредственно при многомерных наблюдениях и измерениях в селекционных испытаниях. Подчеркнуто, что эффективное применение рассмотренных методов возможно лишь при сотрудничестве экспертов предметной области и специалистов в области анализа данных.

**Ключевые слова:** объектно-признаковые данные, анализ формальных понятий, формальный контекст, решетка формальных понятий, импликации на признаках, сельскохозяйственные культуры, пшеница мягкая яровая, засухоустойчивость, устойчивость к бурой ржавчине.

DOI: 10.37313/1990-5378-2020-22-1-86-92

### ВВЕДЕНИЕ

Создание и использование новых сортов занимает ведущее место в решении сложных задач сельскохозяйственного производства. При селекционной работе приходится сталкиваться с анализом данных наблюдений и экспериментов. В этом смысле интересны возможности метода бикластеризации объектно-признаковых данных. Речь идет о такой кластеризации данных, которая лежит в основе онтологического моделирования, когда конструируется структура понятий, описывающая исследуемую предметную область. Бикластер – это пара, описывающая множество объектов и набор свойств, которыми оно обладает.

Для анализа объектно-признаковых данных разработан и математически строго обоснован метод, известный как анализ формальных понятий (АФП) [1, 2]. АФП является одним из самых известных методов кластеризации данных, который получил широкое распространение в самых разных областях и развивается до сих пор [3-5]. Ниже приведена короткая справка о математических основах АФП.

Классический АФП ориентирован на обработку бинарного представления объектно-признаковых данных в виде совокупности оценок истинности семантических суждений вида  $b_{xy}$  = «объект  $x$  обладает свойством  $y$ » и использует следующие обозначения и модели:

•  $K = (G^*, M, I)$  – формальный контекст, где  $G^* = \{g_i\}_{i=1, \dots, r}$ ,  $r = |G^*| \geq 1$  – набор объектов исследуемой предметной области, попавших в поле зрения исследователя (т.е. множество объектов

Семенова Валентина Андреевна, аспирант.  
E-mail: queenbfjr@gmail.com

обучающей выборки:  $G^* \subseteq G$ , где  $G$  – все мыслимое множество объектов предметной области),  $M = \{m_j\}_{j=1, \dots, s}$ ,  $s = |M| \geq 1$  – множество измеряемых у объектов свойств,  $I$  – бинарное соответствие «объекты-свойства», т.е. совокупность оценок  $\|b_{ij}\| \in \{\text{Истина, Ложь}\}$ ;

• операторы Галуа  $\varphi$ ,  $\omega$  (общая нотация « $\cdot$ ») для контекста  $K$ :

$\varphi(X) = X' = \{m_j | m_j \in M, \forall g_i \in X: g_i I m_j\}$  – общие свойства объектов, составляющих  $X \subseteq G^*$ , или Галуа-проекция  $X$  на  $M$ ;

$\omega(Y) = Y' = \{g_i | g_i \in G^*, \forall m_j \in Y: g_i I m_j\}$  – объекты, которые обладают всеми свойствами из  $Y \subseteq M$ , или Галуа-проекция  $Y$  на  $G^*$ ;

•  $(X, Y)$  – формальное понятие, у которого  $X \subseteq G^*$  – объем,  $Y \subseteq M$  – содержание, причем  $X = Y'$ ,  $Y = X'$ ;

•  $B(K)$  – множество формальных понятий контекста  $K$ ;

Согласно АФП и в полном соответствии с классической логикой формальное понятие  $(X_1, Y_1)$  называется *подпонятием* (т.е. является менее общим) другого понятия, или *обобщения*,  $(X_2, Y_2)$ , – обозначается  $(X_1, Y_1) \leq (X_2, Y_2)$ , – если  $X_1 \subseteq X_2$ , или, эквивалентно,  $Y_1 \supseteq Y_2$ . Это отношение порядка задает на множестве всех понятий формального контекста замкнутую решетку  $(B(K); \leq)$ , называемую *решеткой понятий* [1].

Целью данной статьи является демонстрация возможности применения АФП к описанию данных селекционной работы.

## 1. ПОИСК ДАННЫХ ДЛЯ ДЕМОСТРАЦИОННОГО ПРИМЕРА

Для демонстрационного материала мы старались найти данные, которые были бы достаточно систематизированы и представляли бы фундаментальную информацию о селекционной работе. Материал такого качества был найден в каталоге сортов и гибридов сельскохозяйственных культур Поволжского научно-исследовательского института селекции и семеноводства имени П.Н. Константинова [6] и каталоге инновационных разработок Самарского НИИ сельского хозяйства имени Н.М. Тулайкова [7].

В обоих каталогах имеются сведения о результатах селекционной работы в растениеводстве, проводимой в названных институтах. Во втором каталоге также представлен список селекционных сортов, включенных в госреестр, и инновационные технологии возделывания сельскохозяйственных культур. В описание сортов присутствуют стандартные разделы, т.е. информация достаточно хорошо структурирована. В ряде разделов признаки сортов в значительной мере стандартизированы. Однако во многих других содержится неформальная, описательная

информация о сортах. Например, такой характер имеют сведения об урожайности, основных элементах технологии возделывания и т.д.

Тем не менее, для конструирования простейшего примера объектно-признаковых данных, на которых можно будет продемонстрировать возможности анализа формальных понятий, этих сведений вполне достаточно. Конкретно, принято решение воспользоваться данными о сортах мягкой яровой пшеницы из каталога Самарского НИИСХ [7].

## 2. ОТБОР ДАННЫХ ДЛЯ ДЕМОСТРАЦИОННОГО ПРИМЕРА

В качестве объектов выбранной предметной области выступают сорта мягкой яровой пшеницы:

- Жигулевская;
- Самсар;
- Тулайковская 5;
- Тулайковская 10;
- Тулайковская 100;
- Тулайковская золотистая;
- Экада 70;
- Экада 66;
- Тулайковская 110;
- Тулайковская победа;
- Тулайковская надежда;
- Тулайковская 116;
- Экада 214;
- Зауральская волна.

Признаки для демонстрационного примера находим в описании каждого сорта. Напомним, что некоторые характеристики сортов описаны на содержательном уровне, текстуально. Подобная информация трудно формализуема без помощи эксперта в данной предметной области. Так, сведения об урожайности внести в формальный контекст сложно: в каталоге описывается сравнение урожаев в разные годы, в разных регионах, а также иногда отмечаются только максимальные показатели. Однако, формализовать часть некоторых свойств не составило большого труда.

С учетом замечаний выше были избраны свойства, которые можно формализовать неспециалисту: включенность в госреестр, засухоустойчивость, процент содержания белка в зерне и устойчивость к бурой ржавчине.

К сожалению, описание даже этих немногих свойств у отдельных сортов отсутствует. Поэтому недостающие сведения либо добывались в web-сети, либо предполагалось, что сорт имеет примерно такое же значение интересующего свойства, какое имеет его «родитель», указанный в разделе «происхождения». Так, у сорта «Зауральская волна» не отмечено свойство засухоустойчивости, но указывается, что для его возделывания предпочтительна умеренно увлажненная лесостепная зона. По происхождению, данный сорт был получен из гибридной

комбинации Тулайковская 10/Экада 6, где сорт Тулайковская 10 характеризуется высокой засухоустойчивостью, так что совмещение данных обстоятельств наталкивает на вывод, что «Зауральская волна» может быть засухоустойчива. Скорее всего, такой вывод не правомерен в обычных условиях, но для демонстрационного примера такое решение является приемлемым.

Зачастую людей интересует не сам факт наличия свойства у объекта, а именно его значение, или диапазон значений. Именно поэтому в тексте каталога можно прочесть такие формулировки, как «засухоустойчивость», «высокая засухоустойчивость» и «формирует высокие урожаи зерна в экстремальных условиях высоких температур» (что было интерпретировано нами как «экстремально засухоустойчив»). Ясно, что если сорт экстремально засухоустойчив, то он заведомо и высоко засухоустойчив (аналогично с высокой засухоустойчивостью и просто засухоустойчивостью). И в случае, если сорт имеет иммунитет к бурой ржавчине, будем считать, что он также устойчив и достаточно устойчив к бурой ржавчине.

Несколько другая ситуация с характеристикой содержания белка в зерне: здесь наблюдаются изменения в показателях от 12 до 18 % в пределах рассматриваемых сортов. Поскольку для нас не очевидно, на какие диапазоны следует разделять данное свойство, мы обратились

к ГОСТ 9353-2016 для пшеницы [8]. В нем перечислены требования к классам пшеницы, в зависимости от качества зерна, к которому и относится, в том числе, процент белка в зерне.

По сути дела, в ГОСТе осуществлено номинальное концептуальное шкалирование [9, 10], потому что диапазон процента содержания белка важен для определения класса пшеницы в зависимости от качества. Заметим, что для свойств засухоустойчивости и устойчивости к болезням авторами каталога, по сути, было осуществлено порядковое шкалирование [9, 10].

В соответствии с требованиями ГОСТ 9353-2016 нами были сформированы свойства «12-13,5% белка», «13,5-14,5% белка» и «>14,5% белка», которые соответствуют III, II и I классам. В этом смысле, сведения каталога фактически указывают, что зачастую сорт относится к двум классам, хотя в каталоге о принадлежности к классам данных нет.

Ограничимся этим небольшим составом свойств для демонстрационного примера и перейдем теперь к построению таблицы «объекты-свойства».

### 3. КЛАСТЕРИЗАЦИЯ СОРТОВ МЯГКОЙ ЯРОВОЙ ПШЕНИЦЫ

Отобранные данные образуют таблицу «объекты-свойства», т.е. формальный контекст задачи АФП, представленный на рисунке 1. Для

A	B	C	D	E	F
	засухоустойчив	высоко засухоустойчив	экстремально засухоустойчив	12-13,5% белка	13,5-14,5% белка
Жигулевская	X	X			X
Самсар	X				X
Тулайковская 5	X				
Тулайковская 10	X	X			
Тулайковская 100	X	X			X
Тулайковская золотистая	X		X		
Экада 70	X				
Экада 66	X				
Тулайковская 110	X	X			
Тулайковская победа	X	X			X
Тулайковская надежда	X	X		X	X
Тулайковская 116	X	X	X	X	X
Экада 214	X			X	X
Зауральская волна	X				

A	G	H	I	J	K
	>14,5% белка	достаточно устойчив к б.р.	устойчив к б.р.	иммунитет к б.р.	в госреестре
Жигулевская	X	X			X
Самсар	X	X			X
Тулайковская 5	X	X	X	X	X
Тулайковская 10	X	X	X	X	X
Тулайковская 100	X	X	X	X	X
Тулайковская золотистая	X	X	X	X	X
Экада 70	X	X			X
Экада 66	X	X			X
Тулайковская 110	X	X	X		X
Тулайковская победа	X	X	X		X
Тулайковская надежда	X	X	X		X
Тулайковская 116	X	X	X		X
Экада 214	X	X	X		X
Зауральская волна	X	X			X

Рис. 1. Соответствие «сорта-свойства» – формальный контекст демонстрационного примера

каждого объекта-сорта в таблице отмечалось его соответствие каждому из свойств.

Для вывода понятий из подобных объектно-признаковых данных имеются ряд более или менее функционально полных программных средств: ConExp, Galicia, ToscanaJ, Python FCA Tool и др. [11-14]. Воспользуемся одним из самых популярных свободно распространяемым приложением ConExp [11], которое неплохо документировано и имеет интуитивно понятный интерфейс.

Для визуализации выводимой из исходных данных решетки формальных понятий ConExp использует так называемую диаграмму Хассе, т.е. размеченный граф, покрывающий отношение «быть более общим понятием» [15, 16]. Обработка данных демонстрационного примера (рисунок 1) дает диаграмму Хассе, приведенную на рисунке 2.

Узлы данной диаграммы представляют собой формальные понятия со своим объемом и содержанием. Ребра изображают отношение частичного порядка « $\leq$ » на множестве формальных понятий и с точки зрения наследования свойств направлены из каждого узла лишь к

его «соседям сверху». Верхний узел диаграммы имеет максимальный объем, т.е. включает в себя все объекты. В данном случае видно, что всем объектам формального контекста одновременно присущи свойства «засухоустойчив» и «достаточно устойчив к б.р.» («б.р.» – сделанное для удобства сокращение от «бурая ржавчина»). Самый нижний узел содержит все свойства, но не имеет объема – это говорит о том, что ни один объект не обладает сразу всеми свойствами.

Например, согласно рисунку 2 сорт Тулайковская 5 имеет иммунитет к бурой ржавчине, засухоустойчива, характеризуется содержанием белка свыше 14,5% и включена в госреестр. Свойства, характеризующие отдельно взятый узел-понятие, собираем, идя вверх по ребрам, идущим от этого узла, а объекты – идя по ребрам вниз, пример на рисунке 3.

Построенная диаграмма недвусмысленно указывает, например, на то, что засухоустойчивость и устойчивость к бурой ржавчине являются, безусловно, важными свойствами новых сортов, обуславливающими их вывод на рынок.

Помимо решетки понятий методический комплекс АФП позволяет извлекать импликации

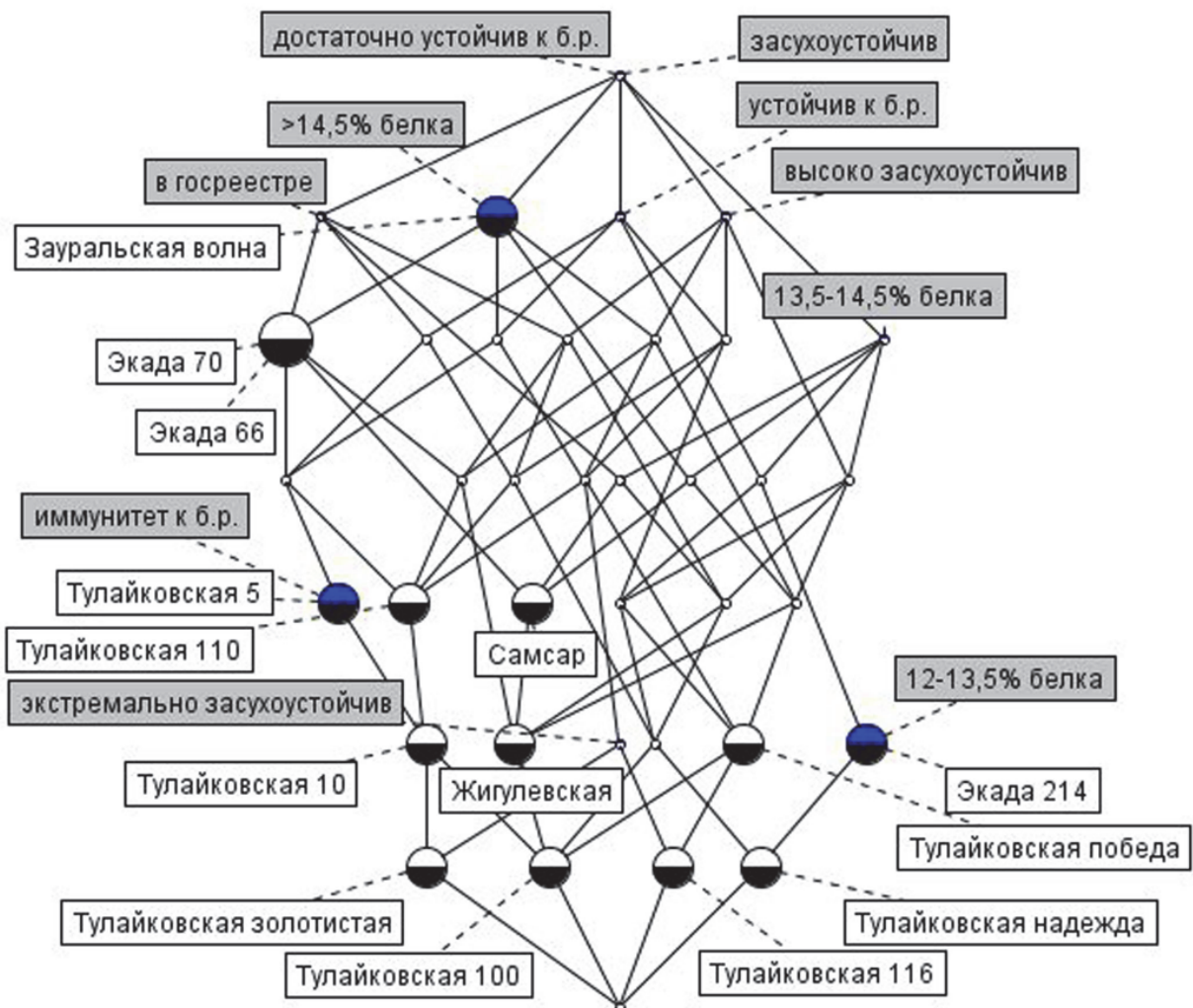


Рис. 2. Диаграмма Хассе, определяющая концептуальное описание сортов мягкой яровой пшеницы

и ассоциативные правила, которым подчиняются признаки объектов. Суть импликации состоит в том, что если объект имеет свойство  $x$ , то он обязательно имеет и свойство  $y$ , но не наоборот. Ассоциативное правило похоже на импликацию, но в этом случае условие «если  $x$ , то  $y$ » выполняется лишь для некоторой части объектов.

Пример одной из импликаций, извлекаемой из рассматриваемого контекста:

«Если сорт засухоустойчив и имеет иммунитет к бурой ржавчине, то он имеет 14,5% белка и включен в госреестр».

В отличие от импликаций количество ассоциативных правил, извлекаемых из нашего небольшого демонстрационного примера очень велико. Для иллюстрации ограничимся одним из них:

«В 86 % случаев если сорт высоко засухоустойчив и устойчив к бурой ржавчине, то он имеет больше 14,5% белка».

Нетрудно заметить, что АФП даже для элементарных случаев, как наш демонстрационный пример, дает весьма большое количество выводимых формальных понятий (см. справку во введении), т.к. метод фиксирует все незначительные

отличия объектов по свойствам. Поэтому методический комплекс АФП включает различные способы редуцирования множества понятий, учитывающие разные критерии («малая поддержка», «высокая чувствительность» и пр.) [17].

Одной из возможностей редуцирования является переход от решетки формальных понятий к специальной таксономии понятий. Такой прием реализован в программном комплексе OntoWorker, разрабатываемом в ИПУСС РАН – СамНЦ РАН [18, 19]. Обработка исходного контекста в упомянутой программе приводит к получению таксономии, представленной на рисунке 4 (изображена транзитивная редукция, когда показываются связи только соседних понятий). К достоинствам такой специальной таксономии понятий относятся более удобное интерпретирование и возможность извлечения неявной, полезной для исследователя информации.

Числами на данном рисунке обозначены виртуальные формальные понятия (т.е. только обобщения реальных объектов-сортов), а числами с приписанной «ф» в начале – фундаментальные понятия, непосредственно представляющие кластеры объектов-сортов [18].

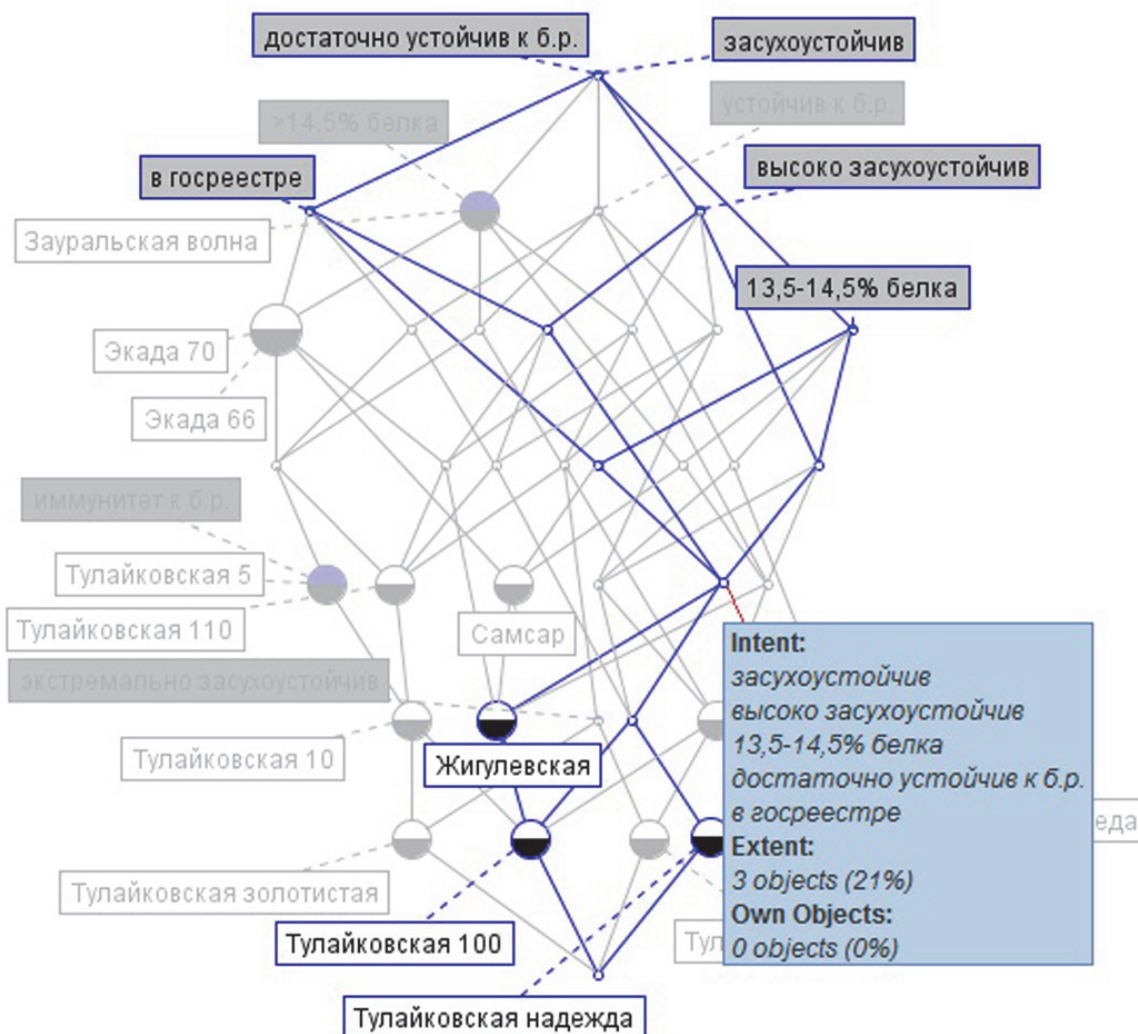


Рис. 3. Содержание и объем формального понятия, демонстрируемые инструментом ConExp

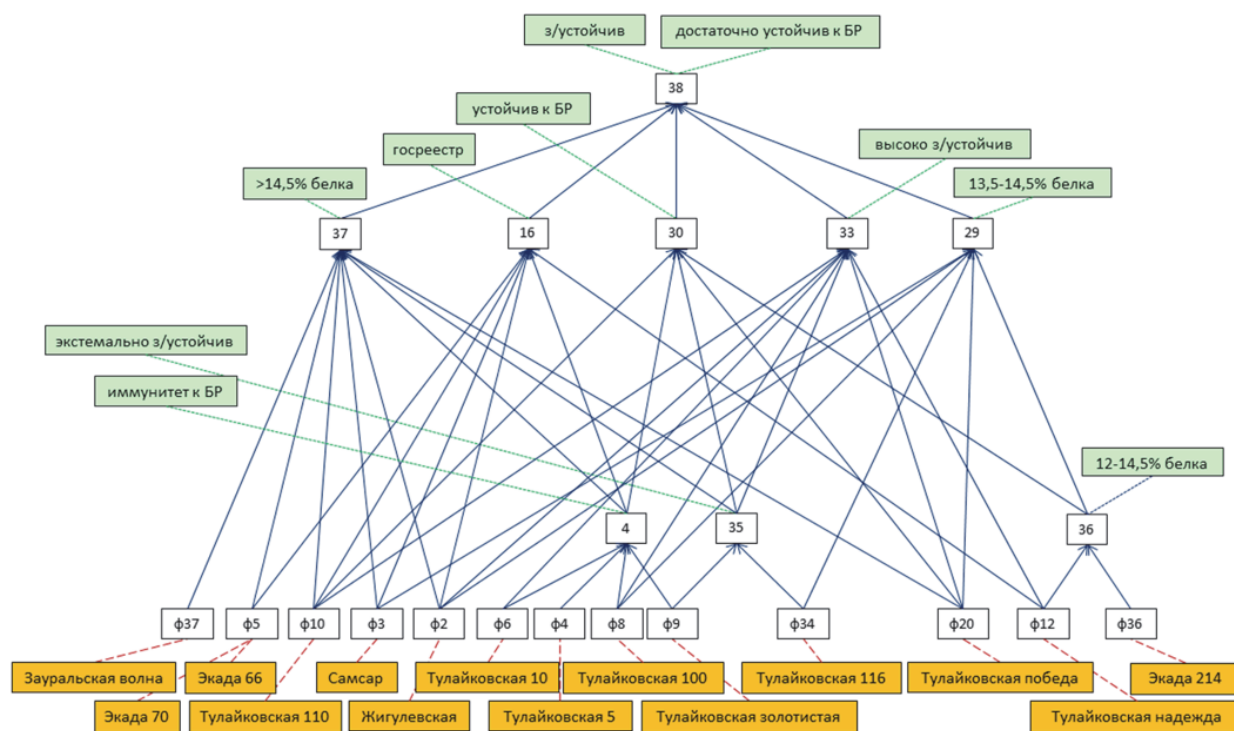


Рис. 4. Специальная таксономия понятий сортов мягкой яровой пшеницы

В такой таксономии концептуальная информация о сортах оказывается ещё более сжатой и представлена в форме, подобной более привычной древовидной иерархии.

### ЗАКЛЮЧЕНИЕ

В статье на демонстрационном примере показана применимость АФП для концептуальной структуризации данных селекционной работы в растениеводстве.

В действительности, входные объектно-признаковые данные, представляющие исследуемую предметную область, могут иметь более богатую структуру. Для каждого объекта сведения могут быть представлены в нескольких строках, отражающих серии измерения. Каждый признак может быть описан несколькими столбцами, представляющими набор конкурирующих процедур. Кроме того, могут учитываться степени достоверности к сериям и степени доверия к процедурам, а также пропуски и отказы от измерений [20]. Упомянутый в статье *OntoWorker* позволяет отражать в формальном контексте все подобные обстоятельства. Очевидно, что такие возможности могут быть при необходимости использованы для обработки промежуточных результатов селекционных исследований.

Разумеется, данные, получаемые с помощью АФП, в состоянии интерпретировать, оценить и использовать только эксперты в исследуемой предметной области, поэтому в общем случае, такой концептуальный анализ данных следует

проводить в сотрудничестве таких экспертов и специалистов по технологиям обработки информации.

### СПИСОК ЛИТЕРАТУРЫ

1. *Ganter B., Wille R.* Formal Concept Analysis. Mathematical foundations. - Berlin-Heidelberg: Springer-Verlag, 1999. 290 p.
2. *Ignatov D.I.* Introduction to Formal Concept Analysis and Its Applications Information Retrieval and Related Fields // In: P. Braslavski, N. Karpov, M. Worrington, Y. Volkovich, D.I. Ignatov (Eds.): Information Retrieval. Revised Selected Papers 8<sup>th</sup> Russian Summer School, 2014 (Nizhniy Novgorod, Russia, August 18-22, 2014). Springer International Publishing, 2015, pp. 42-141.
3. *Смирнов С.В.* Две методологии вывода формальных понятий: когда и как они должны работать вместе // Знания – Онтология – Теория: Материалы VII международной конф. (711 октября 2019 г., Новосибирск, Россия). – Новосибирск: Институт математики СО РАН, 2019. С. 355363.
4. *Пронина В.А., Шипилина Л.Б.* Использование отношений между атрибутами для построения онтологии предметной области // Проблемы управления. 2009. № 1. С.27-32.
5. *Zhang Zh.* Constructing L-fuzzy concept lattices without fuzzy galois closure operation // Fuzzy Sets and Systems. 2017. pp. 71-86.
6. Каталог сортов и гибридов сельскохозяйственных культур селекции ФГБНУ «Поволжский НИИСС» / Под общей ред. В.В. Глуховцева. – Кинель, Поволжский НИИСС, 2016. 61 с.
7. Каталог инновационных разработок Самарского НИИ сельского хозяйства имени Н.М. Тулайкова

- на 2018 год / Под ред. С.Н. Шевченко – Самара: Изд-во СамНЦ РАН, 2018. 92 с.
8. ГОСТ 9353-2016 Пшеница. Технические условия: [Электронный ресурс]. URL: <https://pdf.standartgost.ru/catalog/Data2/1/4293751/4293751950.pdf> (дата обращения: 10.12.2019).
  9. *Ganter B., Wille R. Conceptual scaling // In: F. Roberts (Ed.): Applications of Combinatorics and Graph Theory to the Biological and Social Sciences. – New York Springer-Verlag, 1989. pp. 139-167.*
  10. *Самойлов Д.Е., Смирнов С.В. Субъективные аспекты формирования и обработки данных в анализе формальных понятий // Информационные технологии и нанотехнологии (ИТНТ-2016): Материалы Международной конференции и молодежной школы (17-19 мая 2016 г., Самара, Россия). – Самара: Изд-во СамНЦ РАН, 2016. С. 1038-1043.*
  11. Concept Explorer: [Электронный ресурс]. URL: <http://conexp.sourceforge.net> (Дата обращения: 14.12.2019).
  12. Galicia Lattice Builder: [Электронный ресурс]. URL: <http://www.iro.umontreal.ca/~galicia/> (Дата обращения: 14.12.2019).
  13. ToscanaJ: [Электронный ресурс]. URL: <http://toscanaj.sourceforge.net/> (Дата обращения: 14.12.2019).
  14. Python FCA Tool: [Электронный ресурс]. URL: <https://github.com/ae-hse/fca> (Дата обращения: 14.12.2019).
  15. *Биркгоф Г. Теория решеток. - М.: Наука, 1984. 568 с.*
  16. Диаграмма Хассе – Википедия: [Электронный ресурс]. URL: [https://ru.wikipedia.org/wiki/%D0%94%D0%B8%D0%B0%D0%B3%D1%80%D0%B0%D0%BC%D0%B0\\_%D0%A5%D0%B0%D1%81%D1%81%D0%B5](https://ru.wikipedia.org/wiki/%D0%94%D0%B8%D0%B0%D0%B3%D1%80%D0%B0%D0%BC%D0%B0_%D0%A5%D0%B0%D1%81%D1%81%D0%B5) (Дата обращения: 14.12.2019).
  17. *Kuznetsov S.O. On stability of a formal concept // Ann. Math. Artificial Intelligence. 2007. 49. pp. 101-115.*
  18. *Семенова В.А., Смирнов В.С., Смирнов С.В. OntoWorker: программная лаборатория для онтологического анализа данных // Проблемы управления и моделирования в сложных системах: Труды XVII междунар. конф. (22-25 июня 2015 г., Самара, Россия). – Самара: СамНЦ РАН, 2015. С. 382-393.*
  19. *Лещева Д.В., Семенова В.А., Смирнов С.В. О поддержке когнитивных актов в задачах онтологического анализа данных // Проблемы управления и моделирования в сложных системах: Труды XXI междунар. конф. (3-6 сентября 2019 г. Самара, Россия): в 2-х т. / Под ред.: С.А. Никитова, Д.Е. Быкова, С.Ю. Боровика, Ю.Э. Плешивцевой – Самара: ООО «Офорт», 2019. Т. 2. С. 260-263.*
  20. *Смирнов С.В. Двухсоставность феномена информации и анализ данных (с примерами из когнитивного анализа) // Информационные технологии и нанотехнологии (ИТНТ-2017): Сборник трудов III Международной конференции и молодежной школы (25-27 апреля 2017 г., Самара, Россия). – Самара: Изд-во Новая техника, 2017. С. 1846-1849.*

## EXPERIENCE OF BICLUSTERIZATION APPLICATION TO DATA OF AGRICULTURAL CROPS SORTS

© 2020 V.A. Semenova<sup>1,2</sup>

<sup>1</sup> Institute for the Control of Complex Systems of Russian Academy of Sciences, Samara

<sup>2</sup> Samara State Technical University

The article is devoted to methods for analyzing object-attribute data in low structured areas. Specifically, a demonstrational example of biclusterization of similar data in the field of agricultural crops selection is considered. It is shown what opportunities are appearing for the systematization of empirical data. The data biclusterization method under consideration – a classical formal concept analysis – is an applied branch of lattice theory; a brief reference is given on its mathematical foundations. The source material for the demonstrational example was borrowed from the catalog of innovative developments of the Samara Research Institute of Agricultural Sciences named after N.M. Tulaykov. Sorts of soft spring wheat were chosen as objects, and several properties which are describing them were chosen as attributes. A number of these properties are scaled in accordance with state standard 9353-2016 for wheat. Based on the selected and adjusted material, the initial table “objects-properties”, or a formal context for the formal concept analysis, is formed. To process this data, we used the free software package ConExp, which implements a methodical complex for formal concept analysis. The main results that these methods give for selection data structuring are demonstrated. These results include a lattice of formal concepts, implications and associative rules on a set of attributes. Moreover, the result of the OntoWorker software package being developed at ICCS RAS – SamSC RAS is demonstrated. The outcome consists in transforming and reducing the lattice of formal concepts into a special taxonomy of formal concepts (classes) that is more convenient for the user to perceive and interpret. It is noted that advanced methods of formal concept analysis can take into account the incompleteness of the considered data, determined by the series multiplicity of measurement experiments with different degrees of results reliability, and by the presence of competing measurement procedures with different degrees of confidence in their results. This reflects the realities of the empirical information accumulation directly during multidimensional observations and measurements in selection probations. It was emphasized that the effective application of the considered methods is possible only in cooperation of knowledge data expert and specialists in the field of formal concept analysis. *Keywords:* object-attribute data, formal concept analysis, formal context, lattice of formal concepts, implication on attributes, agricultural crops, soft spring wheat, drought resistance, brown rust resistance. DOI: 10.37313/1990-5378-2020-22-1-86-92