

# КЛАСТЕРИЗАЦИЯ ЦИФРОВЫХ ПЛАТФОРМ ДЛЯ УПРАВЛЕНИЯ ЭКОНОМИКОЙ СТРАНЫ

**Ерешко Ф.И., Мединников В.И.**

*Вычислительный центр им. А.А. Дородницына, ФИЦ ИУ РАН,  
Россия, г. Москва, ул. Вавилова, д.44-2  
fereshko@yandex.ru, dommed@mail.ru,*

**Богатырева Л.В.**

*Институт проблем управления им. В.А. Трапезникова РАН,  
Россия, г. Москва, ул. Профсоюзная д.65  
lbogat@mail.ru*

*Аннотация: в работе рассматривается математическая модель кластеризация цифровых платформ для управления экономикой страны. Дается оригинальное определение цифровой платформы для снятия неопределенности в связи с множественными трактовками данного понятия.*

Ключевые слова: кластеризация, цифровая платформа, математическая модель.

## **Введение**

Понятие цифровой платформы связано с появлением Программы цифровой экономики России, в которой ставится цель создания не менее 10 цифровых платформ (ЦП). Поскольку при этом не были четко определены понятия и критерии их формирования, то без такого понятийного аппарата началась широкая трактовка их, что ведет к запутыванию смысла цифровизации экономики. В научной литературе встречаются десятки определений, исходя из рода деятельности и достигнутых результатов субъектов, дающих собственную трактовку.

Суммируя все эти факты, можно констатировать, что под ЦП многие в стране понимают некую внешнюю площадку (платформу), обеспечивающую работу организаций в сфере новых форм платежей и коммуникаций с потребителями, но никак не новые формы управления и экономические отношения. Такая широкая трактовка этого понятия ведет к запутыванию смысла цифровизации экономики и несет большую угрозу достижению целей Программы цифровой экономики. Например, в сельском хозяйстве, опираясь на такое положение с определением, некоторые исследователи пишут о ЦП выращивания некоторой культуры в заданном регионе на конкретной геоинформационной системе. В таком случае их окажется около 5 млн. шт. [1]. Этот же принцип закладывается и в соответствующие предложения по ЦЭ АПК, что не способствует интеграции информационных ресурсов и информационных систем в единое целое.

## **1 Понятие цифровой платформы**

Мы считаем, что научно-обоснованный подход к выполнению с наибольшей эффективностью Программы цифровой экономики возможен только при четком осмыслении понятия ЦП на основе системного анализа процесса формирования этих платформ и его формализации на основе математического моделирования. Не найдя в литературе формализованного описания формирования ЦП, дадим такое описание на основе опыта разработки автоматизированных систем управления предприятиями и организациями на разных уровнях управления в стране. Будем придерживаться следующего определения.

Цифровая платформа управления экономикой – совокупность упорядоченных цифровых данных на основе онтологического моделирования; математических алгоритмов, методов и моделей их обработки и программно-технических средств сбора, хранения, обработки и передачи данных и знаний, оптимально интегрированных в единую информационно-управляющую систему, предназначенную для управления целевой предметной областью с организацией рационального цифрового взаимодействия заинтересованных субъектов.

## **2 Модель формирования цифровых платформ**

Под структурой системы управления будем понимать организационную совокупность ее взаимосвязанных элементов, определяющих их место как в чисто физическом, так и технологическом смысле.

Под проектированием структуры цифровых платформ понимается процесс построения взаимосвязей элементов структуры управления и самих элементов в соответствии с заданными критериями эффективности в целом.

Рассматривается система, состоящая из множества узлов управления  $j$  (например, федеральных и региональных министерств, ведомств, предприятий, их подразделений), множества задач  $K$ ,

связанных с обработкой данных, размещаемых в дата-центрах, ситуационных центрах (СЦ), кластеров данных  $L$ , типов связи  $R$ . Процесс управления предполагается периодически с периодом  $T$ , и все операции расчетов, передачи данных и т.д. усреднены по времени. Будем считать, что любая задача может решаться в любом узле, в том числе разбиваться по этим узлам. Для решения задач используются некоторые обобщенные технические средства.

Математическая модель

$k$  - номер задачи,  $k \in K$ ;  $l$  - номер информационного элемента,  $l \in L$ ;  $j$  - номер узла управления,  $j \in J$ ;  $f_{klj}^e$  - средние характеристики (объем информации; временные, частотные требования и т.д.) на информацию  $l$ -ого элемента, необходимого для задачи  $k$ , возникающий в узле  $j$ ,  $e \in E$ ;  $x_{jk} = 1$ , если  $k$ -я задача решается в узле  $j$ , 0 – иначе;

$\alpha_{klj} = 1$ , если  $l$ -й элемент возникает в узле  $j$  для  $k$ -й задачи, 0 – иначе;

$y_{lj_1j_2r} = 1$ , если  $l$ -й элемент передается из  $j_1$ -го узла в  $j_2$ -й посредством  $r$ -го средства связи;

$d_{mjk}$  - необходимые ресурсы  $m$ -го типа для решения  $k$ -й задачи в  $j$ -м узле;

$M_m$  -  $m$ -е ресурсы оборудования;

$s_{lj_1j_2r} = 1$ , если  $r$ -й тип связи используется для передачи  $l$ -го элемента из  $j_1$ -го узла в  $j_2$ -й;

$G_r^e$  - характеристики средств связи;  $c_j^1$  - стоимость единицы оборудования в  $j$ -м узле;  $c_{j_1j_2r}^2$  - стоимость  $r$ -го средства связи при передаче информации из  $j_1$  в  $j_2$ ;  $c_{j_1j_2r}^3$  - затраты на передачу единицы информации из  $j_1$  в  $j_2$ ;  $c_{mjk}^4$  - стоимость  $m$ -го ресурса для решения  $k$ -й задачи в  $j$ -м узле;  $c_k^5$  - обобщенная стоимость  $k$ -й задачи;  $c^0$  - средства, выделенные на разработку ЦП;

Ограничения на размещение задач по узлам и техническим средствам:

$$(1) \sum_j x_{jk} \geq 1, \quad k \in K^3 \in K, \text{ то есть } k\text{-я задача должна быть решена хотя бы в одном узле};$$

(2)  $x_{jk} \geq 1, \quad j \in J_1, \quad k \in K^4 \in K$ , т.е. некоторые задачи из множества  $K$  должны быть обязательно решены в некоторых узлах  $j \in J_1$ ;

Условия передачи информации из узла  $j_1$  в узел  $j_2$ :

$$(3) \sum_r y_{lj_1j_2r} = \sum_k \alpha_{klj_1} x_{j_1k}, \quad j_1 \neq j_2.$$

Информация передается из узла  $j_1$  в узел  $j_2$ , когда она возникает в узле  $j_1$  и используется в узле  $j_2$  для задачи  $k$ ;

$$(4) \sum_r y_{lj_1j_2r} \leq 1, \quad \text{информация передается одним средством связи.}$$

Ограничение на загрузку оборудования:

$$(5) \sum_{jk} d_{mjk} x_{jk} \leq M_m.$$

Ограничения на каналы связи:

$$(6) \sum_{l,k} y_{lj_1j_2r} f_{klj_2}^e \leq G_r^e s_{j_1j_2r}.$$

Финансовые ограничения на инвестиции:

$$(7) \sum_{j,k} c_j^1 x_{jk} + \sum_{j_1, j_2, r} c_{j_1j_2r}^2 s_{j_1j_2r} + \sum_{j,k} c_k^5 x_{jk} \leq c^0.$$

Критерий эффективности:

$$(8) \sum_{j,k} c_j^1 x_{jk} + \sum_{j_1, j_2, r} c_{j_1j_2r}^2 s_{j_1j_2r} + \sum_{j_1, j_2, r} c_{j_1j_2r}^3 f_{klj_2}^e y_{lj_1j_2r} + \sum_{m,j,k} c_{mjk}^4 d_{mjk} x_{jk} + \sum_{j,k} c_k^5 x_{jk} \rightarrow \min$$

Представленная в работе модель распределяет в пределах выделенных финансовых ресурсов информационные средства и решаемые задачи по узлам управления (дата-центрам, СЦ), определяет при

необходимости инвестиции в телекоммуникационные средства с оптимизацией информационных потоков. В сформированных группах информационных ресурсов возможна ситуация, когда отдельные информационные массивы, сгруппированные подобным образом, окажутся несвязанными или слабо связанными между собой, что не соответствует данному выше определению ЦП. Для дальнейшего определения ЦП воспользуемся кластерным анализом, применяемым для кластеризации предметных областей пользователей при проектировании баз данных, например [2]. При априори заданном количестве (кластеров) ЦП в [3] дан строгий математический алгоритм определения степени общности предметных областей пользователей на основе меры подобия, применяемой в теории автоматической классификации. Для классификации ЦП можно воспользоваться предлагаемой функцией подобия. В нашем же случае число кластеров неизвестно заранее. Тогда можно воспользоваться алгоритмами, приведенными в [4]. В последнее время в связи с появлением интернета огромным интересом стали пользоваться методы кластеризации текстовых документов. Один из популярных методов основан на теории графов кластеризации и построении минимального остовного дерева по алгоритму Краскала [5, 6].

Используя один из представленных методов, кластеризацию ЦП в каждом дата-центре можно провести на основании так называемой матрицы семантической смежности на полученных значениях  $f_{klj}^e$  для конкретного дата-центра  $j_0$ . Элемент  $a_{in}$  (коэффициент сходства) матрицы семантической смежности  $\|a_{in}\|$  представляет собой величину в диапазоне от 0 до 1, равную количеству пересечений элемента  $i$  с элементом  $n$  во всех решаемых задачах (можно учесть разные характеристики групп, например, частоту, объем, важность использования информации и т.д.), отнесенных к количеству элементов (количеству пересечений и т.д.). Если две группы обладают большим сходством, то они должны принадлежать одной ЦП, если их сходство равно нулю либо меньше некоторого порогового критического значения, то они должны быть в разных ЦП.

Одним из результатов кластеризации ЦП является научно обоснованный расчет потребности в необходимых специалистах для ЦЭ, который в правительственных кругах опирается на аналогичные западные данные и устаревшие технологии так называемого позадачного подхода без учета интеграции информационных ресурсов и информационных систем путем формирования стандартов на их представление, функции управления, как это предлагалось в проекте ОГАС [7]. Опишем основные группы специалистов в каждой конкретной ЦП (отрасли). Во-первых, это онтологи. Во-вторых, специалисты в области создания баз данных. В-третьих, специалисты в области разработки архитектуры больших информационных и информационно-управляющих систем. В-четвертых, программисты, умеющие разрабатывать большие информационные и информационно-управляющие системы с различными режимами обработки информации. В-пятых, специалисты в области информационной безопасности. И, конечно, самая большая группа – специалисты по внедрению и сопровождению ИС и ИУС. Каждая из указанных групп состоит из различных отраслевых и специфических подгрупп, количество исполнителей в которых зависит от количества задач в ЦП.

## Литература

1. Ерешко Ф.И., Кульба В.В., Меденников В.И. Интеграция цифровой платформы АПК с цифровыми платформами смежных отраслей // АПК: экономика, управление, 2018, № 10. – С. 34-46.
2. Дж. Мартин. Планирование развития автоматизированных систем. М.: Финансы и статистика, 1984.
3. Кульба В.В., Микрин Е.А., Павлов Б.В., Платонов В.Н. Теоретические основы проектирования информационно-управляющих систем космических аппаратов. М.: Наука, 2006.
4. Afifi A. H., Clark V. Computer Aided Multivariate Analysis. London: Chapman & Hall, 1996. – P. 412.
5. Белоусов А. И., Ткачев С. Б. Дискретная математика. — М.: МГТУ, 2006. – P. 744.
6. Joseph. B. Kruskal. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. // Proc. AMS. 1956. Vol 7, No. 1. – P. 48-50.
7. Меденников В.И. Единое информационное Интернет-пространство АПК на основе идей А.И. Китова и В.М. Глушкова об ОГАС. // Цифровая экономика, 2018, №1. – С. 38-49.