

Реализация баз данных типа "ключ-значение" клеточными автоматами с локаторами

Э. Э. Гасанов¹, А. А. Пропажин²

В данной работе показано, что базы данных типа “ключ-значение” могут быть реализованы клеточными автоматами с локаторами таким образом, что время выполнения основных операций, таких как поиск, вставка, удаление, не будет зависеть от размера базы данных и будет равно суммарной длине ключа и значения.

Ключевые слова: Клеточные автоматы с локаторами, базы данных, ключ-значение.

База данных “ключ-значение” — это популярная сейчас парадигма хранения данных, также называемая словарем. Такую базу данных можно представлять в виде множества пар строк (k, v) , где первая строка k называется ключом и служит идентификатором пары, а вторая строка v называется значением. *Строка* — это последовательность символов некоторого алфавита A , оканчивающаяся специальным символом \emptyset , называемым *символом окончания строки*, причем символ \emptyset не принадлежит алфавиту A .

База данных “ключ-значение” поддерживает следующие операции:

1) *вставка пары* (k, v) — в базе данных появляется запись с ключом k и значением v ; если запись с ключом k уже имела в базе данных, то значение заменяется на v ;

2) *удаление записи с ключом* k — из базы данных удаляется запись (k, v) ; если записи с ключом k в базе данных нет, то база данных не изменяется;

3) *поиск элемента по ключу* k — в базе данных находится запись (k, v) , и значение v возвращается в качестве ответа; если записи с ключом k в базе данных нет, то ответом служит пустое множество.

Понятие клеточного автомата с локаторами было введено Э. Э. Гасановым в работе [1] и уточнено Г. В. Калачевым в работе [2]. Точное определение можно найти в упомянутых работах, здесь же мы приведем

¹Гасанов Эльяр Эльдарович — профессор каф. математической теории интеллектуальных систем мех.-мат. ф-та МГУ, e-mail: el_gasnov@mail.ru.

Gasnov Elyar Eldarovich — professor, Lomonosov Moscow State University, Faculty of Mechanics and Mathematics, Chair of Mathematical Theory of Intelligent Systems.

²Пропажин Артем Алексеевич — студент каф. математической теории интеллектуальных систем мех.-мат. ф-та МГУ, e-mail: artem.propazhin@mail.ru.

Propazhin Artem Alekseevich — student, Lomonosov Moscow State University, Faculty of Mechanics and Mathematics, Chair of Mathematical Theory of Intelligent Systems.

неформальное определение одномерного клеточного автомата с одним полным локатором, который и будет использоваться в данной работе.

Одномерный клеточный автомат представляет собой множество одинаковых элементарных автоматов, расположенных в целочисленных узлах вещественной прямой, и называемых клетками. Поведение клеточного автомата задается функцией переходов, а именно состояние клетки в следующий момент однозначно определяется собственным состоянием в текущий момент и состояниями его соседей. Далее будем считать, что у каждой клетки ровно два соседа: ближайший слева и ближайший справа. Одно из состояний клетки называется состоянием покоя, и если клетка и ее соседи находятся в состоянии покоя, то и в следующий момент клетка останется в состоянии покоя. В клеточном автомате с одним полным локатором кроме алфавита состояний имеется алфавит вещания с заданной на нем коммутативной полугрупповой операцией. В данной работе в качестве такой операции будет использоваться операция максимум. В клеточном автомате с одним полным локатором каждая клетка каждый момент посылает в эфир некоторый сигнал из алфавита вещания, определяемый функцией вещания. Значение функции вещания клетки определяется собственным состоянием клетки, состояниями ее соседей и суммарным сигналом, полученным из эфира. Суммарный сигнал эфира формируется суммированием с помощью полугрупповой операции сигналов вещания всех клеток за исключением сигнала данной клетки. В клеточном автомате с одним полным локатором значение функции переходов клетки также определяется собственным состоянием клетки, состояниями ее соседей и суммарным сигналом, полученным из эфира.

Введем еще одну сущность — *пользователя* базы данных. Будем считать, что пользователь базы данных имеет возможность посылать в эфир сигналы из алфавита вещания и получать из эфира сигналы из алфавита вещания. Будем считать, что клеточный автомат с локаторами вместе с пользователем реализуют базу данных типа “ключ-значение”, если алфавит вещания представляет собой множество, состоящее из пар вида (“команда”, $A \cup \{0\}$), где “команда” — принимает одно из значений: “поиск”, “вставка”, “удаление”, “ответ”, “нет ответа”, “нет команды” (здесь “нет команды” является минимальным элементом), и поведение клеточного автомата задается следующим образом.

1) Пользователь подает в эфир команду “поиск” и первый символ ключа k , а затем последовательно все остальные символы ключа, включая символ 0. Если в базе данных нет записи с ключом k , то на следующий такт после подачи символа 0 клеточный автомат выдает в эфир команду “нет ответа”. Если в базе данных есть запись (k, v) , то на следующий такт после подачи символа 0 клеточный автомат выдает в эфир

пару (“ответ“, a), где $a \in A$ — первый элемент значения v , и в последующие такты клеточный автомат последовательно выдает в эфир все остальные символы значения v вплоть до символа 0 включительно.

2) Пользователь подает в эфир команду “вставка“ и первый символ ключа k , а потом последовательно все остальные символы ключа, включая 0, и после этого таким же способом передаются символы значения v . В результате в базе данных, реализуемой клеточным автоматом, появляется пара (k, v) , т.е. если в последующем подать на вход клеточному автомату команду “поиск“ и ключ k , то клеточный автомат вернет значение v в ответ.

3) Пользователь подает команду “удаление“ и первый символ ключа k , а потом последовательно все остальные символы ключа, включая 0. В результате из базы данных, реализуемой клеточным автоматом, исчезает запись с ключом k , т.е. при последующем поиске по ключу k клеточный автомат вернет “нет ответа“.

Теорема 1. *Существует клеточный автомат с локаторами и пользователем, который реализует базу данных типа “ключ-значение“, и для которого время выполнения операций поиск, вставка, удаление не будет превышать суммарную длину ключа и значения.*

Приведем идею доказательства данной теоремы.

Будем использовать одномерный клеточный автомат с одним полным локатором, причем элементарные автоматы, лежащие в отрицательной области числовой прямой использоваться не будут. Выше уже был описан алфавит вещания. Алфавит состояний состоит из троек вида (“команда“, “тип ячейки“, $A \cup \{0, *\}$), где “команда“ принимает одно из значений: “поиск“, “вставка“, “удаление“, “ответ“; “тип ячейки“ принимает одно из значений: “командир“, “текущая ячейка типа ключ“, “текущая ячейка типа значение“, “необрабатываемая ячейка“, “приемник для записи“; * — специальный символ. Символом “приемник для записи“ будет помечаться элементарный автомат, начиная с которого будет осуществляться следующая запись в базу данных. В начальный момент этим символом будет помечен автомат с номером 0. Записи базы данных будут представлять собой пары строк ключ-значение, причем первый символ каждой записи будет отмечен состоянием “командир“.

Опишем функционирование данного автомата. Когда начинается вставка, все командиры получают из эфира сигнал в виде (“вставка“, k_1), где k_1 — первый символ ключа. Если k_1 совпадает со значением, хранящимся в состоянии клетки командира, то в состоянии следующей справа клетки команда меняется на “вставка“, а тип ячейки на “текущая ячейка типа ключ“. На следующий такт следующая справа от командира клетка, также получив сигнал из эфира, проверяет совпадение с хранящимся

символом. При совпадении происходит аналогичный предыдущему процесс за тем исключением, что в конце такта тип этой клетки изменится на “необрабатываемая ячейка”. Процесс происходит последовательно до символа 0 включительно. Процесс проверки ключа начинается одновременно со всех командиров. Если в какой-то момент случилось несовпадение, то тип следующей клетки меняется на “необрабатываемая ячейка”. Если дошли до символа 0, то это означает, что мы нашли запись с искомым ключом, и эту запись надо исключить из базы данных. Исключения из базы данных достигается тем, что 0 в состоянии текущей клетки меняется на символ *. В результате при последующих поисках совпадения не будет.

Одновременно с поиском ключа, начиная с клетки, помеченной состоянием “приемник для записи”, происходит последовательная запись пары “ключ-значение” в базу данных. В момент прихода команды “вставка” клетка, находящаяся в состоянии “приемник для записи”, становится командиром и сохраняет в своем состоянии первый символ ключа k_1 . При этом в следующий момент клетка, находящаяся правее данной клетки меняет команду на “вставка” и тип ячейки на “приемник для записи”. Далее клетки, имеющие команду “вставка” и тип ячейки “приемник для записи” сохраняют в своем состоянии поступающие из эфира символы ключа и значения. При этом состояние (“вставка”, “приемник для записи”) переходит к следующей справа клетке. Исключением является момент прихода символа 0 для значения. В этот момент следующая справа клетка переходит в состояние (“поиск”, “приемник для записи”).

Удаление происходит аналогично вставке. Только во время удаления не начинается процесс записи в конец базы данных.

Во время поиска считывание ключа происходит так же, как и при вставке. Но дойдя до символа 0, замена данного символа на * не происходит. Клетка, следующая справа от клетки с символом 0, меняет команду на “поиск” и тип ячейки на “текущая обрабатывающая ячейка типа значение”. Клетки, находящиеся в этом состоянии, отправляют в эфир сигнал (“ответ”, a), где $a \in A$ — хранимый символ. Текущая клетка после отправки сигнала меняет тип на “необрабатываемая ячейка”. Сигналы с символами значения последовательно отправляются в эфир до сигнала с символом 0 включительно.

Список литературы

- [1] Гасанов Э. Э., “Клеточные автоматы с локаторами”, *Интеллектуальные системы. Теория и приложения.*, **24:2** (2020), 121–133.
- [2] Калачев Г. В., “Замечания к определению клеточного автомата с локаторами”, *Интеллектуальные системы. Теория и приложения.*, **24:4** (2020), 47–56.

Implementation of key-value databases by cellular automata with locators

E.E.Gasanov, A.A.Propazhin

In this paper, it is shown that key-value databases can be implemented by cellular automata with locators in such a way that the execution time of basic operations, such as search, insert, delete, will not depend on the size of the database and will be equal to the total length of the key and value.

Keywords: Cellular automata with locators, key-value databases.

References

- [1] Gasanov E. E., “Cellular automata with locators”, *Intelligent Systems. Theory and applications*, **24**:2 (2020), 121–133 (In Russian).
- [2] Kalachev G. V., “Remarks on the definition of a cellular automaton with locators”, *Intelligent Systems. Theory and applications*, **24**:4 (2020), 47–56 (In Russian).