

Автоматизация проверки карт знаний (mind maps)

Алексей Г. Дубинский^[0000–0001–8536–1603]

ГУ «Днепропетровская медицинская академия МОЗ Украины»,
ул. В. Вернадского, 9, Днепр, 49044, Украина
dubinsky@ukr.net

Аннотация. Карты знаний, которые составляют студенты, подлежат проверке. Сравниваем их с образцом, который подготовил лектор. Для этого переходим к моделям карт — орграфам типа дерева. Для сравнения графов требуется разработать приложение, подключаемое к системе управления обучением на сервере. Чтобы уменьшить сложность разработки приложения нужно ограничить выбор текстов, добавляемых студентами на карту знаний, выбором терминов из тезауруса предметной области.

Ключевые слова: карта знаний, автоматизация проверки, математическая модель интеллект-карты, программное обеспечение.

1 Введение

Метод диаграмм связей, карт знаний или интеллект-карт (mind maps) уже несколько десятилетий пытаются внедрить в образовании. В целом, наблюдается полезный эффект, величина которого зависит от целого ряда факторов [1]. Мы полагаем, что ключевым фактором есть выбор между активным и пассивным подходом. Пассивный — когда ученики получают от учителя готовые карты, как еще один вид иллюстративного материала. Активный — когда студенты составляют свои mind maps в ходе обучения. В [2] мы говорим про синтез этих двух подходов с целью соединить их достоинства и компенсировать недостатки.

Каждую карту знаний, которая создана студентами, нужно проверить и оценить. Простое произведение количества студентов в потоке соответствующего курса на количество тем, предусмотренных в учебном плане дисциплины, дает нам количество карт, которые преподаватель должен рассмотреть. Получаем величину порядка нескольких тысяч, если речь идет об общем курсе, который изучают все студенты, проходящие обучение по данной специальности. К сожалению, академическая нагрузка лектора не предусматривает

дополнительное оплаченное рабочее время для проверки такого количества работ студентов.

Казалось бы, проверку работ студентов можно поручить преподавателям-ассистентам, ответственным за практические занятия. Однако это создает проблему с оцениванием, когда разные преподаватели могут ставить неодинаковые баллы за студенческие работы сопоставимого качества. К тому же будет намного сложнее выявлять факты академической недобросовестности, когда один студент копирует готовую работу другого студента и сдает копию как свою чтобы получить оценку. Также нет возможности изыскать дополнительное оплаченное рабочее время преподавателей.

Другой путь — совместное оценивание — часто используют в массовых открытых онлайн курсах [3] (МООС). Опыт работы в отечественных вузах свидетельствует о кооперативных, а не состязательных установках подавляющего большинства студентов, потому организовать эффективное совместное оценивание весьма нетривиальная задача.

Таким образом мы приходим к необходимости автоматизации проверки работ студентов. Нужен метод определения качества составленных студентами карт знаний по конкретной теме учебного курса. Этот метод должен быть несложным, чтобы реализующее его программное обеспечение могло быть разработано в местном вузе, например, студентами, специализирующимися в сфере инженерии программного обеспечения, в ходе выполнения их дипломного проекта. Важное требование к методу — объяснительная сила, чтобы студенту можно было аргументированно пояснить, почему одни работы получает более высокую оценку чем другие.

2 Сравнение двух интеллект-карт

Для сравнения двух интеллект-карт нам потребуется перейти от графического изображения (на бумаге или на экране компьютера) к какой-то формальной модели. Будем рассматривать только компьютерные интеллект-карты, ибо для обработки работ на бумаге требуется решать задачи сканирования, векторизации полученного растрового изображения, распознавания текста и связей между узлами и т. д.

Любая интеллект-карта (mind map) складывается в первую очередь из множества текстовых элементов (терминов, концептов), которые соединены между собой с помощью линий (прямых или кривых).

Узлы карты могут быть дополнены изображениями и комментариями. Текстовые элементы могут отличаться размером, цветом, гарнитурой и другими параметрами шрифта. Соединительные линии могут иметь разный тип, цвет и толщину. Линии могут быть основными и необязательными (дополнительными).

Представляется очевидным, что наиболее удобной математической моделью интеллект-карты является граф [4], вершинам которого приписаны текстовые элементы (термины предметной области), а дуги графа представляют связи между ними.

Типичная интеллект-карта может быть представлена как ориентированный граф типа дерева, где выделен корень — вершина, которая имеет нулевую степень захода. Это центральный элемент карты, основная тема, которой посвящена интеллект-карта. Все дуги графа направлены, во все остальные вершины заходит только одна дуга.

Мы полагаем, что в первом приближении при сравнении можно пренебречь такими элементами интеллект-карт как иллюстративные изображения, комментарии к дугам (в частности, указывающим на их семантику, например, IsA/MemberOf, AKindOf/SubsetOf, HasPart) и т. п. Ибо важно не столько тип иерархии, который автор счел важным изобразить на диаграмме, сколько сам факт выбора данных терминов/концептов.

Что касается изображений, следует учитывать, что добросовестный преподаватель при составлении интеллект-карт обычно старается соблюдать законодательство об авторских правах. То есть он может добавлять лишь те изображения, использование которых в явном виде разрешено владельцами авторских прав, например, с помощью свободных лицензий Creative Commons.

Студент же при выполнении задания свободен от такого ограничения. И, следовательно, может использовать более релевантные изображения, которые защищены копирайтом. Поэтому сравнивать изображения, размещенные на разных интеллект-картах, не столько трудно, сколько бессмысленно и нецелесообразно.

Таким образом задача сравнения двух интеллект карт сводится к задаче сравнения двух ориентированных графов, вершинам которых соответствуют текстовые строки. Это короткие тексты, обычно состоящие из нескольких слов, обычно являющиеся терминами изучаемой предметной области.

Как сравнить такие графы? Начнем со сравнения параметров этих графов, без учета текстов, приписанных к вершинам. Сначала

посмотрим на количество вершин графа, глубину дерева (наибольшую длину пути от корня к листу) и фактор сбалансированности дерева (разность наибольшей и наименьшей длин путей от корня к листам).

Допустим мы хотим сравнить интеллект-карту к лекции про компьютерные сети из [2, fig. 4] с простыми картами [5, рис. 1] и [5, рис. 2]. Результаты подсчета даны в таблице 1.

Таблица 1. Пример сравнения графов трех карт

Источник	[2, fig. 4]	[5, рис. 1]	[5, рис. 2]
Порядок графа	49	14	22
Глубина дерева	3	4	2
Фактор сбалансированности	1	2	0

Видно, что у первого графа значительно больше вершин, у второго больше глубина, а третье дерево сбалансировано. Очевидно, этих фактов недостаточно для оценки сходства графов. Можно лишь заключить, что второй граф недостаточно полно описывает предметную область, т. к. количество его вершин составляет менее трети числа вершин первого графа. Сравнив порядок графа можно предположить, что третья карта также менее информативна чем первая.

Задача сравнения графов достаточно хорошо изучена, см., например, обзоры [6, 7]. Различия двух графов можно определить через расстояние редактирования, которое вычисляется как минимальное количество элементарных операций редактирования узлов (удаление, вставка, переименование), которое преобразует одно дерево в другое [8]. Очевидный недостаток этого подхода — не учитывается вес (важность) редактируемых узлов.

В нашем случае порядок графов мал, потому нет необходимости выбирать методы сравнения, оптимизированные по скорости выполнения или по объему используемой памяти. Конкретизируя задачу, укажем, что сравнивать будем деревья с помеченными вершинами. Следовательно, до запуска процедуры оценки изоморфизма графов необходимо решить задачу определения соответствия помеченных вершин этих графов.

3 Определение тождества узлов карты (вершин графа)

Рассуждая в терминах теории графов, мы концентрируем внимание на связях между понятиями (концептами, записями), представленными в интеллект-картах. Но в реальности эти понятия (концепты, записи), важнее связей между ними. Следовательно, необходимо в первую очередь определить какие именно понятия (концепты, записи) использовал студент в своей интеллект-карте.

Здесь возникает ряд трудностей, связанных с распознаванием семантики текстовых строк. Перечислим некоторые причины их возникновения:

- Слабое владение терминологией предметной области, когда студент вместо нужного термина использует его определение или иной термин;
- Использование слов-синонимов;
- Синтаксические ошибки (особенно характерны для иностранных студентов);
- Объединение нескольких терминов (узлов графа, карты) вместе в одну запись.

Таким образом получаем достаточно сложную задачу, у которой нет простых решений. Особенno, если мы намерены работать с потоками студентов, у которых отличаются языки обучения.

Чтобы избежать необходимости распознавать смысл текста, введем ограничения на вводимый студентами текст. Вместо прямого ввода с клавиатуры будем предлагать студенту сделать выбор из списка, содержащего ограниченное количество вариантов.

Вариантами выбора будут термины предметной области, которые, по мнению преподавателя релевантны рассматриваемой теме. Задача автоматизации составления терминологического словаря рассмотрена ранее в [9]. Этот способ использовался нами при подготовке терминологических словарей по четырем основным предметам, которые преподаются на кафедре медико-биологической физики и информатики Днепропетровской медицинской академии.

В словаре термины пронумерованы и каждому из них соответствует число — порядковый номер термина в словаре. Теперь для меток вершин наших графов вместо текстовой строки можно использовать эти числа. Если терминологические словари составлены отдельно по

каждой теме, то индекс, указывающий на конкретный термин, будет состоять из пары чисел: (порядковый номер темы; номер термина в словаре темы).

Таблица 2. Пример: фрагмент пронумерованного списка терминов по курсу «медицинская информатика»

96	добавить линию тренда
97	доверительная вероятность
98	доверительный интервал
99	доказательная медицина
100	достоверность коэффициента корреляции
...	
297	размер поля
298	размер шрифта
299	разметка страницы
300	раскрывающийся список
301	распределение
302	расстояние между строками
303	регрессионный анализ

4 Оценка веса (важности) терминов

Как мы уже отмечали, оценка различия двух графов через расстояние редактирования не позволяет напрямую учитывать важность редактируемых узлов. Совершенно очевидно, что ошибка выбора термина для узла верхнего уровня должна быть оценена как более существенная по сравнению с ошибкой выбора термина для вершин нижнего уровня (листов дерева).

Присвоим листьям нашего орграфа (концевые вершины, самый нижний уровень детализации интеллект-карты) минимальный вес, равный единице. Определить вес вершин следующего уровня можно несколькими способами. Во-первых, взять наибольшую длину пути от выбранной вершины к листу, тогда вершины следующих уровней получат веса 2, 3, 4 и т. д. тогда одна ошибка с вершиной третьего уровня будет эквивалентна трем ошибкам в вершинах- листьях.

Другой способ — рассчитывать вес кумулятивно, исходя из количества листьев, собранных в данной ветви дерева. На рис. 1 подсчитаны такие веса для фрагмента интеллект-карты из [2, fig. 4].

Без учета свернутых листьев, которые скрыты на изображении, вершина «based on size» получила вес 4, «Networks Classifications» — вес 13.

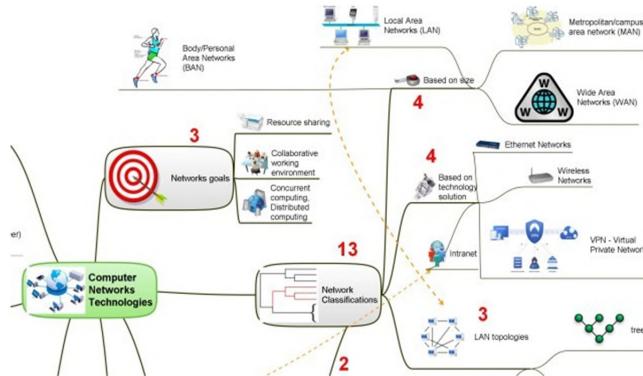


Рис. 1. Фрагмент интеллект-карты по теме «компьютерные сети» [2, fig. 4]

Все вершины данного фрагмента карты, получившие вес больше 1 перечислены в таблице 3.

Если учитывать не только листья, но и все промежуточные вершины (такие как «based on size»), тогда «Networks Classifications» получит вес 17.

Таблица 3. Список вершин фрагмента интеллект-карты

Узел интеллект-карты	Вес
Network Classifications	13
Based on size	4
Based on technology solution	4
LAN topologies	3
Networks goals	3
Data transmissions communication channel	2

5 Сравнение на основе весов вершин

В части 2 мы говорили о сравнении двух интеллект- карт с помощью сравнения двух ориентированных графов, без учета важности

(весов) вершин. Рассмотрим теперь способ сравнения на основе весов помеченных вершин, без учета возможных различий их соединений в графе.

Каждый пронумерованный термин предметной области может быть размещен на исследуемой интеллект-карте. Подсчитаем его вес, исходя из количества листьев, собранных в данной ветви дерева, как указано выше. Термины, не размещенные на интеллект карте, получают вес 0. Таким образом вся нужная нам информация будет записана как массив целых чисел a_i , где индекс массива i есть номер термина в списке (терминологическом словаре), а значение a_i — подсчитанный вес этого термина.

Такая модель содержания документов хорошо известна и активно используется в информационном поиске [10]. Наш массив весов есть вектор, характеризующий содержание данного документа (в нашем случае — интеллект-карты). Различие двух документов определено как расстояние в многомерном векторном пространстве. В простейшем случае используем евклидову метрику. То есть расстояние между двумя документами можно легко подсчитать как корень квадратный из сумм квадратов разностей соответствующих значений из двух массивов.

$$d(\mathbf{a}, \mathbf{b}) = \sqrt{\sum_i (a_i - b_i)^2} \quad (1)$$

Далее можно перейти к обратной величине и нормировать полученный результат к интервалу $[0;1]$, чтобы говорить не про расстояние между векторами, а про их близость (сходство).

6 Требования к программному обеспечению

Программное обеспечение, которое будет выполнять проверку интеллект-карт должно запускаться на сервере. Необходима совместимость с системой управления обучения Moodle, подключение к ней как модуль/ плагин, чтобы получить доступ к интеллект- картам студентам, которые они будут загружать в виде файла или составлять в режиме онлайн. Для работы с загружаемыми файлами потребуется возможность распознавать стандартные форматы файлов наиболее популярных программ создания mind map.

Результат сравнения программа должна выдавать как в качественном виде (оценка) так и детально, в количественном, как результат сравнения графов. Оценка должна быть доступна как в

традиционной шкале оценивания знаний — от 2 до 5. Итоговая оценка за весь курс нужна также и в шкале ECTS (от А до F).

7 Выводы

Автоматическая оценка качества интеллект-карт, созданных студентами при освоении тем учебного курса может быть выполнена указанным способом. Вводимые ограничения, позволяющие решить эту задачу, заключаются в запрете на ввод текстов с клавиатуры. Подписи к узлам карты должны быть выбраны из списка, подготовленного заранее, — из терминологического словаря предметной области.

Необходимость ограничения текстового ввода требует отказаться от использования распространенных программ и онлайн-сервисов построения интеллект-карт. Фактически нужно разработать новое программное обеспечение или же модифицировать один из существующих программных пакетов, доступных по open source

лицензии. Возможно, для первых итераций следует ограничиться только текстовым представлением карты в режиме outline, как это реализовано в сервисе text2mindmap.

Литература

1. Liu, Y., Zhao, G., Ma, G., Bo, Y. (2014). The effect of mind mapping on teaching and learning: A meta-analysis. Standard Journal of Education and Essay, 2 (1), 17–31.
2. Dubinsky, A. (2019). Mind maps for education. In: Exploring the Mind's Eye: An Interdisciplinary Conference on Imagination, October 25–26, 2019, Bilkent University, Bilkent, Ankara, Turkey. URL: <http://repo.dma.dp.ua/4447/>
3. Datcun N.N. (2015). Sovmestnoe otsenivanie deiatelnosti obuchaiushchikhsia v massovyykh otkrytykh onlain kursakh: sistematicheskii obzor literatury. Mir nauki, (3), 10–10. URL: <https://mir-nauki.com/PDF/53PDMN315.pdf>
4. Siochos, Vasilis and Papatheodorou, Christos Developing a Formal Model for Mind Maps., 2011. In First Workshop on Digital Information Management, Corfu (Greece), 30–31 March 2011. [Conference paper] URL: <http://eprints.rclis.org/15842/>
5. Dubinskii A. Ispolzovanie diagramm sviazey (mindmaps) v lektcionnykh prezentaciiakh. // Suchasni tendentsii navchannia pryrodnycho-matematychnykh ta tekhnolohichnykh dystsyplin u zahalnoosvitnii

- ta vyschii shkoli: [materialy III Mizhnarodnoi naukovo-praktychnoi onlain- internet konferentsii, m. Kropyvnytskyi, 17–22 zhovtnia 2016 roku] / Kropyvnytskyi: RVV KDPU im. V. Vynnychenka, 2016. — 176 s.
6. Sanfeliu, A., Fu, K. S. (1983). A distance measure between attributed relational graphs for pattern recognition. IEEE transactions on systems, man, and cybernetics, (3), 353–362. doi: 10.1109/TSMC.1983.6313167
 7. Bunke, H. (2000, September). Recent developments in graph matching. In Proceedings 15th International Conference on Pattern Recognition. ICPR-2000 (Vol. 2, pp. 117–124). IEEE. doi: 10.1109/ICPR.2000.906030
 8. Rachkovskii D. A. Bystryi poisk skhodnykh grafov po rasstoianiu redaktirovaniia // Kibernetika i sistemnyi analiz — 2019, Tom 55 — № 6, s. 178–194.
 9. Dubinskii A. G. Avtomatizatsiia podgotovki uchebnogo terminologicheskogo slovaria // Mizhnarodna naukovo-praktychna konferentsiya «Informatsiini tekhnolohii v osviti ta nautsi» Melitopol, MDPU, 13–14 chervnia 2019 r.
 10. Dubinskii A. G. Nekotorye voprosy primeneniia vektornoi modeli predstavleniia dokumentov v informacionnom poiske // Upravliaiushchie sistemy i mashiny. — 2001. — T. 2. — S. 77.

Automation of mind maps

Olexyi H. Dubinsky [0000-0001-8536-1603]

State Institution “Dnipropetrovsk Medical Academy of the Ministry of Health of Ukraine”, 9, Vernadsky str., Dnipro, 49044, Ukraine

dubinsky@ukr.net

Abstract. The mind maps that students compile are subject to verification. To compare graphs, you need to develop an application that connects to the control system training on the server. To reduce the complexity of application development, you need to limit the choice of texts added by students to the knowledge map, the choice of terms from the thesaurus subject area.

Automatic assessment of the quality of intelligence maps created by students when mastering the topics of the course can be performed in this way. The restrictions imposed to solve this problem are the prohibition of entering texts from the keyboard. Signatures to card nodes must be selected from a list prepared in advance — from terminological dictionary of the subject area.

The need to limit text input requires abandoning the use of common programs and online services for building smart maps. In fact, it is necessary develop new software or modify one of the existing software packages available under an open source license. Perhaps, for the first iterations it is necessary to be limited only to text representation of a card in the outline mode as it is implemented in text2mindmap service.

Keywords: knowledge map, verification automation, mathematical model of intelligence map, software.