

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АНГАРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ИНСТИТУТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ГЕОФИЗИКИ СО РАН

**НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
В ИССЛЕДОВАНИИ
СЛОЖНЫХ СТРУКТУР**

**МАТЕРИАЛЫ
ТРИНАДЦАТОЙ МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ
7–9 сентября 2020 г.**

Томск
Издательский Дом Томского государственного университета
2020

отдельное облако. С другой стороны, в реальных задачах для визуализации большого числа объектов чаще всего используются полигональные модели. Таким образом, возникает задача перехода от облачного представления к полигональному, которая имеет множество вариантов решения, обладающих, однако, значительной вычислительной сложностью и невозможностью работы в реальном времени. Однако если предположить, что имеется некоторая база полигональных моделей эталонных объектов, которые с высокой вероятностью могут встретиться во входной сцене, возникает альтернативный способ: вместо построения полигональной модели по облаку точек можно найти соответствующую готовую модель, если таковая есть в базе, и использовать ее для визуализации объекта в соответствующей 3D-среде. Для корректного сравнения объектов-запросов с объектами-эталоном в базе хранятся не только полигональные модели, но и преобразованные к облачному виду.

Вместо прямого сравнения входного объекта со всеми объектами из базы предлагается использовать сравнение дескрипторов их формы, среди которых можно выделить как инвариантные к аффинным преобразованиям, так и неинвариантные, обычно более точные, но при этом напрямую неприменимые, если сравниваемые модели имеют разный масштаб или по-разному сориентированы в пространстве [1]. Таким образом, дополнительно встает задача «правильной» ориентации всех моделей перед сравнением, для решения которой предлагается найти собственные значения и собственные векторы ковариационной матрицы C

координат точек, описывающих объект: $C = \frac{1}{N} \sum_{i=1}^N (\mathbf{P}_i - \bar{\mathbf{P}})(\mathbf{P}_i - \bar{\mathbf{P}})^T$. Собственные векторы ковариационной

матрицы C определяют новый базис в пространстве, в котором объект наиболее «вытянут» вдоль первого собственного вектора. Второй собственный вектор ортогонален первому и определяет наибольшую «вытянутость» (разброс значений) точек объекта среди всех возможных направлений, ортогональных первому собственному вектору. Третий вектор ортогонален первым двум. Таким образом, собственные векторы можно использовать для составления матрицы поворота в пространстве, который необходимо выполнить для всех моделей перед построением дескрипторов, подобных GASD, SPIN [2, 3]. Однако найденные векторы определяют лишь оси, вдоль которых располагаются главные компоненты, но при этом направления векторов (знаки) остаются не определены. Для того чтобы избежать перебора вариантов поворота, предлагается методика, которая позволяет определять нужные направления главных компонент (знаки) независимо от начального положения. В ее основе лежит нахождение дополнительных характеристик (свойств) частей исходного облака точек, лежащих по разные стороны плоскости, проведенной через центр масс объекта и перпендикулярной оси, соответствующей главной компоненте.

Кроме того, собственные значения ковариационной матрицы C сами по себе дают первичное представление о некоторых характеристиках формы рассматриваемого объекта и позволяют производить первичное «грубое» сравнение. Сравнивать объект целесообразно только с теми объектами из базы, которые имеют схожий набор собственных значений. Однако нужно соблюдать осторожность в ситуациях, когда некоторая пара из тройки собственных значений или вся тройка собственных значений оказываются близкими друг к другу. Это означает, что в случае представления этого же объекта в виде облака точек после другого сканирования очередность главных компонент может измениться, и объект будет повернут иначе. В подобных случаях лучше воспользоваться дескрипторами, инвариантными к повороту модели.

Литература

1. Приступа А.В., Лапатин И.Л., Замятин А.В. Применение глобальных дескрипторов формы для идентификации 3D объектов // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2019. № 48. С. 57–66. DOI: 10.17223/19988605/48/7
2. Do Monte Lima J.P.S., V. Teichrieb. An efficient global point cloud descriptor for object recognition and pose estimation // Proceedings of the 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images. Sao Paulo, Brazil, 4–7 October 2016. P. 56–63.
3. Johnson A.E., M. Hebert. Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes // IEEE Trans. Pattern Analysis and Machine Intelligence. 1999. Vol. 21(5). P. 433–449.

РАСШИРЕНИЕ СЛОВАРЯ РНРМОРНУ ДИАЛЕКТНЫМИ СЛОВАМИ*

М.Л. Громов, С.С. Земичева

Национальный исследовательский Томский государственный университет, Томск, Россия
{maxim.leo.gromov, optysmith}@gmail.com

В настоящее время в Лаборатории общей и сибирской лексикографии (ЛОСЛ) Томского государственного университета ведутся исследования народно-речевой культуры Среднего Приобья. Основным

* Исследование выполнено при финансовой поддержке гранта Российского научного фонда, проект № 19-78-10015 «Разработка электронных ресурсов для исследования народно-речевой культуры Среднего Приобья».

материалом этих исследований являются тексты, записанные в экспедициях по региону на протяжении 70 лет [1], и диалектные словари, составленные по результатам этих экспедиций [2–4]. Одним из масштабных проектов Лаборатории, осуществляемым в настоящее время под руководством С.С. Земичевой, является оцифровка диалектного архива и создание на его основе электронного лингвистического корпуса [5]. Одной из важнейших операций с корпусом текстов является поиск заданного слова. В принципе, эта задача давно и успешно решается. Для этого достаточно построить инверсный индекс корпуса текстов (для каждого слова составить список текстов, в которые это слово входит). Далее задача становится тривиальной, необходимо сформулировать нужный запрос к базе данных. Однако такой поиск для проводимых в ЛОСЛ исследований недостаточен. Требуется также находить тексты, в которых заданное слово входит в какой-либо своей форме (например, искать не только слово «мама», но и «мамы», «маме» и т.д.). Кроме того необходимо искать слова и пары слов с заданной грамматической информацией. Например, найти прилагательное женского рода в родительном падеже. Причём может быть указана только грамматическая информация без указания самого слова (то есть найти все такие слова). Данная задача тоже разрешима. Среди инструментов, которые позволяют работать с грамматической информацией слов, есть свободно распространяемый инструмент phpMorphy [6]. Он позволяет по заданному слову находить все его формы, в том числе и базовые, хотя и с некоторой грамматической омонимией (для слова «мыла» будут найдены как минимум две базовые формы: существительное «мыло» и глагол «мыть»).

Однако, phpMorphy не учитывает диалектное происхождение текстов корпуса. Точнее, это не учтено в его внутреннем словаре, там, естественно, нет диалектных слов Среднего Приобья. Поэтому возникла необходимость дополнить словарь phpMorphy диалектными словами.

Данная работа посвящена разработке инструмента для облегчения процесса добавления новых слов в словарь phpMorphy. Разработка была разбита на несколько этапов. На первом этапе необходимо выяснить, какие диалектные слова уже есть в словаре phpMorphy, какие он правильно угадывает (то есть находит правильную базовую форму слова), а каких в нём нет (и не угадываются). Далее был разработан инструмент по внесению отсутствующих слов в словарь phpMorphy [7].

Литература

1. Лаборатория общей и сибирской лексикографии. Экспедиции. URL: <http://losl.tsu.ru/?q=node/4>.
2. Палагина В.В. (ред.) Словарь русских старожильческих говоров средней части бассейна р. Оби. Томск: Томский государственный университет, 1964–1967. Т. 1–3.
3. Блинова О.И., Палагина В.В. (ред.) Словарь русских старожильческих говоров средней части бассейна р. Оби. Дополнение. Томск: Томский государственный университет, 1975. Т. 1–2.
4. Палагина В.В. (ред.) Среднеобский словарь: Дополнение. Томск: Томский государственный университет, 1975. Т. 1–2. С. 1983–1986.
5. Zemcheva S.S., Ivantsova E.V. The project of Tomsk dialect corpus in keeping with trends of corpus linguistics development // Siberian Journal of Philology. 2018. № 3. P. 192–205. DOI: 10.17223/18137083/64/18 <https://elibrary.ru/item.asp?id=35597565>
6. phpMorphy. URL: <http://phpmorphy.sourceforge.net>.
7. M.Gromov Add Lemma phpMorphy. URL: https://github.com/maxim-leo-gromov/add_lemma_phpMorphy.