

ИНФОРМАТИКА И КОМПЮТЪРНИ НАУКИ INFORMATICS AND COMPUTER SCIENCES

ИЗСЛЕДВАНЕ ВЪЗМОЖНОСТИТЕ НА ГЕОКОДИРАЩАТА СИСТЕМА НЗ, РАЗРАБОТЕНА ОТ UBER TECHNOLOGIES

Павел Петров

Икономически университет – Варна

Резюме: *Изследването се фокусира върху една прилагана в практиката геокодираща система с почти равноплощна теселация, която използва като база за разделение на сферичната земна повърхност фигурите петъгълник и шестоъгълник. Получената в резултат на делението почти равноплощна теселация предоставя интересни възможности за бързо определяне на равно отдалечени разстояния и визуализиране на т.нар. „топлинни карти“. Последното дава възможност лесно да се анализират динамични и бързо променящи се данни по отношение разпределението на измерваните стойности на единица площ. В тази връзка са очертани най-съществените характеристики на системата НЗ. Тази система може да се използва ефективно при географски информационни системи, свързани с управление на транспортни средства и потоци. Последното се подпомага от множество програмни библиотеки, които се разпространяват с лиценз отворен изходен код. Изследването може да послужи за основа при избора на подходяща геокодираща система при нужда от анализ и визуализация на динамични, бързо променящи се данни, асоциирани с географско местоположение.*

Ключови думи: *НЗ, Uber Technologies, геокодиране, програмни библиотеки, равноплощна теселация*

ВЪВЕДЕНИЕ

Бързата обработка на големи обеми пространствени данни в днешно време е една от приоритетните области на развитие и системите за геокодиране са от съществено значение за анализиране на големи обеми от геопространствени данни (Vasilev et al. 2022). Съществуват множество различни геокодиращи системи, всяка от които разделя пространството на подпространства с различна геометрична форма – най-често на триъгълник, четириъгълник или шестоъгълник. Използват се различни системи за създаването на уникалния идентификатор – код на пространството или геокод (Petrov et al. 2018). Допълнително системите се различават и при: дълбочина на рекурсивното подразделяне на пространствата; характеристики на пространствата – размери, площ; на колко подпространства директно се дели едно пространство и др.

Всички тези разлики дефинират и различните възможности по отношение на съхраняване, обработка и предаване на геокода, както и удобството при агрегиране на данни, статистическа обработка и анализ, визуализация и моделиране. Разделянето на пространството на равноъгълни по размери подпространства – клетки, на база равенство на

ъглите на географската ширина и географската дължина, използва сравнително елементарни аритметични операции. В резултат на равноъгловото подразделяне отделните новообразувани клетки са с различни размери в мерни единици за дължина/разстояние и едновременно с това – с различни площи.

Във връзка с това цел на настоящото изследване е да се проучи същността на геокодиращата система НЗ, предлагана от компанията Uber Technologies, и да се очертаят ползите от нейното прилагане в географски информационни системи.

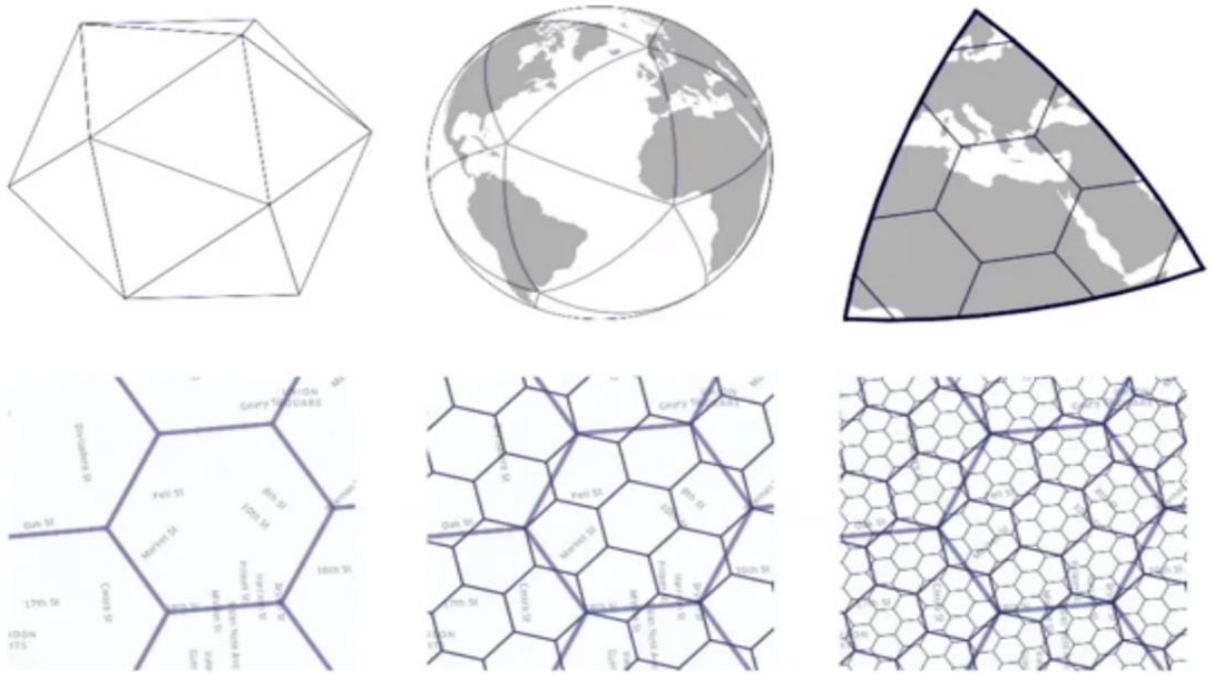
МЕТОДОЛОГИЯ НА ИЗСЛЕДВАНЕТО

При направеното проучване са изследвани литературни и онлайн източници от големи ИТ компании, предлагащи и използващи активно в своята дейност геокодиращата система с равноплощна теселация НЗ. Събраната информация е анализирана и систематизирана със задачата да се синтезират най-важните и характерни моменти на тази използвана в практиката геокодираща система с равноплощна теселация. Получените данни могат да се използват при прилагане на геокодиращи системи с равноплощна теселация, като се използват идеите, добрите практики и положителните характеристики на съществуващите системи. Правим уточнението, че използването на многоъгълници с пет и шест ъгъла за равноплощна теселация може да се прилага, но задължително следва да се има предвид, че площите са само приблизително равни, а не точно равни между отделните фигури на едно и също ниво на резолюция.

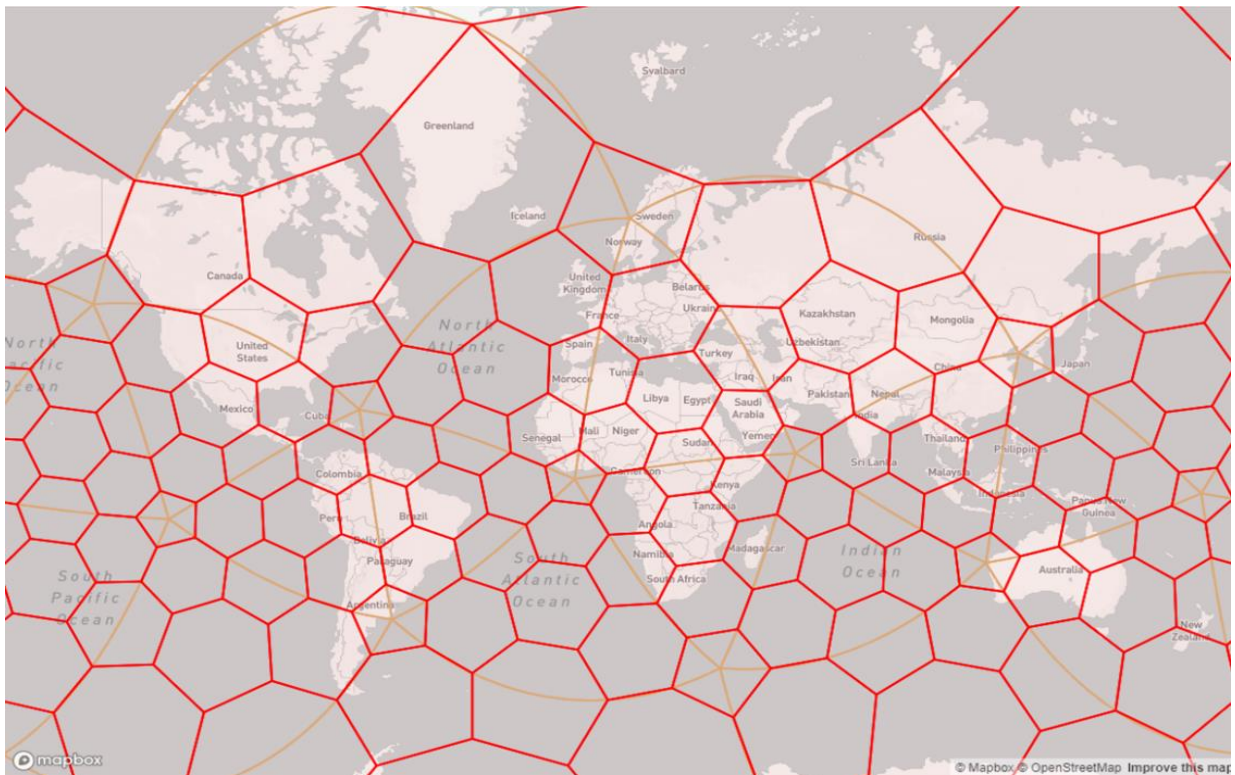
РЕЗУЛТАТИ

НЗ е геокодираща система, разработена от Uber Technologies, при която пространството се разделя на шестоъгълници и петоъгълници на йерархичен принцип. Шестоъгълните подпространства от едно и също ниво на резолюция имат приблизително еднакви площи, а само петоъгълните – еднакви площи (Uber Engineering 2019, Bondaruk et al. 2020, Macdonald et al. 2022).

На фиг. 1 и фиг. 2 е показано първоначално разделяне на карта в меркаторова проекция на шестоъгълници и петоъгълници с приблизително еднакви площи. Виждат се ръбовете на икосаедрона (геометрично тяло с 20 страни), използван при теселацията. Икосаедронът е така ориентиран, че 12-те му върха, където изкривяванията са най-големи, да не са на суша, а на вода. При първоначалното разделяне (ниво на резолюция 0) има 122 пространства – 110 шестоъгълника (хексагона) и 12 петоъгълника (пентагона). На пространствата се задават номера от 0 до 121.



Фиг. 1. Основни геометрични моменти при разделяне на пространствата при системата H3
(Източник: Uber Engineering 2019)



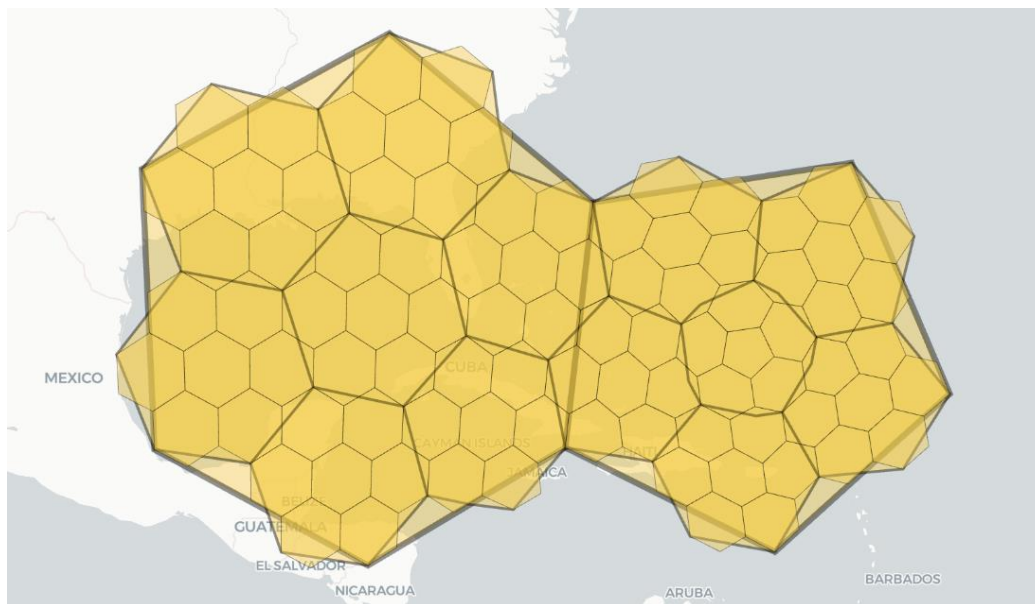
Фиг. 2. Първоначално разделяне на карта в меркаторова проекция чрез геокодиращата система H3
(Източник: H3 Index Inspector – <https://observablehq.com/@nrabinowitz/h3-index-inspector>)

При последващо деление родителският шестоъгълник се разделя на 7 шестоъгълни пространства. Както се вижда на фиг. 3, части от вложените шестоъгълници излизат извън обхвата на родителския шестоъгълник, а части от други съседни шестоъгълници попадат в обхвата на родителския шестоъгълник. Площта на частите, които излизат, и площта на частите, които влизат в рамките на родителския шестоъгълник, са равни.



Фиг. 3. Разделяне на родителски шестоъгълник на седем вложени шестоъгълника (Източник: H3 Documentation 2022a)

Докато шестоъгълниците се подразделят на 7 шестоъгълника, то петоъгълниците се подразделят на 6 пространства – 5 шестоъгълника и 1 петоъгълник (фиг. 4).

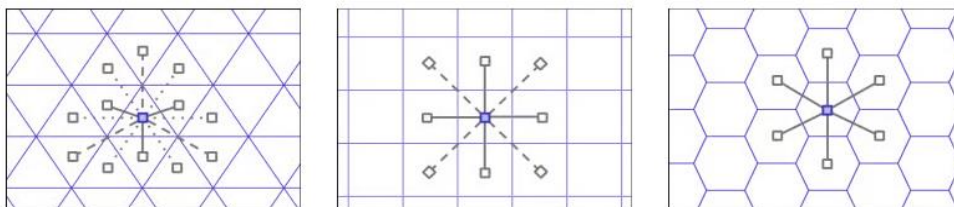


Фиг. 4. Схема на подразделяне на шестоъгълници и петоъгълници на подпространства (Източник: H3 Documentation 2022b)

Предимството на шестоъгълниците спрямо триъгълна или квадратна решетка е, че няма съществена разлика между вида на съседите на една централна клетка – с всичките съседни има обща страна (ръб). При квадратите има два различни вида съседни: при първия вид се

споделя обща страна (в хоризонтална и вертикална посока), а при другия вид се споделя връх (в диагоналните посоки).

Поради тази причина системата НЗ се счита за подходяща за бизнес целите на Uber Technologies, тъй като разстоянията между централен шестоъгълник и неговите шест съседни при шестоъгълна решетка са постоянни, докато при триъгълна решетка имаме 12 съседа и 3 различни разстояния, а при квадратна – 9 съседа и 2 различни разстояния (фиг. 5).



Фиг. 5. Разстояние между клетка и нейните съседни при триъгълна, квадратна и шестоъгълна решетка (Източник: Brodsky 2018)

От бизнес гледна точка информационната система на Uber Technologies, използваща НЗ, може да се разглежда като пазар (marketplace) на транспортни услуги за отделни квартали, градове и държави, където фундаменталните за икономическата теория сили на търсене и предлагане (в случая на транспортни услуги) формират пазара. В този смисъл може да се каже, че информационната система обслужва множество относително слабо свързани помежду си регионални автомобилни транспортни пазари на ниво град или държава (Calafiore et al. 2021; Xia 2022).

Както е известно, за да функционира оптимално един свободен пазар, е нужно свободно да циркулира информация, необходима на участниците на пазара (търсещи и предлагачи) обосновано да вземат своите решения. Част от тази информация може да е в резултат на преобразуване на „сурови“, първични данни чрез изследване, анализ и подходящо визуализиране. На фиг. 6 е показан процес на създаване на „топлинна карта“, използвана при анализ на пазара на транспортни услуги в даден район. „Топлинната карта“ е начин за визуализация на данни, при който чрез палитра от различни цветове се представят числови стойности, отнасящи се за определено пространство.

Подобна карта е възможно да послужи да се вземе решение за по-оптимално разпределение на ресурсите в конкретен момент (Jamil 2020; Min et al. 2021).



Фиг. 6. Процес на създаване на „топлинна карта“ за брой превозни средства в клетка, използвана при анализ на пазара по отношение на търсене и предлагане на транспортни услуги при НЗ (Източник: Brodsky 2018)

Геокодът, наречен H3Index, представлява 64-битово число без знак, което се представя като 15- или 16-символен низ – числото, записано в шестнадесетична бройна система. Първите битове на числото съдържат служебна информация, а в следващите битове, групирани по тройки, са записани номерата на клетките. Например H3 геокод „8a1ec09a2447fff“ е на територията на УниБИТ – виж фиг. 7. На фигурата това е централната клетка, която има следните параметри: ниво на резолюция 10, координати (г.ш. 42.65768413068898, г.д. 23.38775378812013), площ 16478,4 кв.м, дължина на страната 79,6 м.



Фиг. 7. Използване на системата H3 на Uber Technologies при геокодиране на местоположение на територията на УниБИТ (Източник: <https://observablehq.com/@nrabinowitz/h3-index-inspector#8a1ec09a2447fff>)

Исходен код на програмна библиотека за работа с H3 с лиценз свободен софтуер е наличен в платформата GitHub (<https://github.com/uber/h3-js>). Класът разполага с методи, чиито имена са самоописателни и поради тази причина в повечето случаи смисълът им се подразбира. По наше мнение методите на класа могат да се разделят в три основни категории в зависимост от значението им: базови методи, спомагателни методи и методи с универсално приложение.

ИЗВОДИ/ДИСКУСИЯ

Основният извод, който можем да направим на база на изложеното, е, че значителна част от геокодиращите системи, създавани и използвани от големите ИТ компании, са равнопложни. Поради тази причина считаме за перспективно съществуващата система H3 да се използва, тъй като по този начин значително се улеснява сравняването и анализирането на данни, асоциирани с различни клетки.

ЗАКЛЮЧЕНИЕ

В заключение можем да обобщим, че използването на шестоъгълници улеснява приблизителното изчисляване на разстоянието между клетките на решетката и поради това е подходящо за определени задачи, свързани както с определяне на маршрути, така и с транспорта като цяло. При H3 се комбинират предимствата на шестоъгълна глобална решетъчна система с йерархична система за индексирание.

Благодарности: Това изследване е финансирано по проект НПИ 67-2023 от ИУ – Варна.

REFERENCES

- Bondaruk, B., S. Roberts, C. Robertson.** (2020). Assessing the state of the art in Discrete Global Grid Systems: OGC criteria and present functionality. *Geomatica*, 74(1), 9–30. DOI: 10.1139/geomat-2019-001.
- Brodsky, I.** (2018). H3: Uber’s Hexagonal Hierarchical Spatial Index. [viewed 10 October 2023]. Available from: <https://www.uber.com/blog/h3/>.
- Calafiore, A., N. Murage, A. Nasuto, F. Rowe.** (2021). Deriving Spatio-temporal geographies of human mobility from GPS traces. In *Spatial Data Science Symposium 2021 Short Paper Proceedings*, 1–9. DOI: 10.25436/E26K5F.
- H3 Documentation** (2022a). *Indexing*. [viewed 10 October 2023]. Available from: <https://h3geo.org/docs/highlights/indexing>.
- H3 Documentation** (2022b). *Appendix: Methodology*. [viewed 10 October 2023]. Available from: <https://h3geo.org/docs/core-library/restable>.
- Jamil, R.** (2020). Uber and the making of an Algoticon-Insights from the daily life of Montreal drivers. *Capital & Class*, 44(2), 241 – 260. DOI: 10.1177/0309816820904031.
- Macdonald, J., L. Dolega, A. Singleton.** (2022). An open source delineation and hierarchical classification of UK retail agglomerations. *Scientific Data*, 9(1), Article number: 541. DOI: 10.1038/s41597-022-01556-3.
- Min, S., K. So, M. Jeong.** (2021). Consumer adoption of the Uber mobile application: Insights from diffusion of innovation theory and technology acceptance model. In *Future of tourism marketing*, 2–15.
- Petrov, P., P. Dimitrov, S. Petrova.** (2018). Geohash-EAS – a Modified Geohash Geocoding System with Equal-Area Spaces. SGEM 18 International Multidisciplinary Scientific Geoconference Conference Proceedings, Vol. 18 Informatics, Geoinformatics and Remote Sensing, № 2.2, 187–194. DOI: 10.5593/sgem2018/2.2.
- Uber Engineering** (2019), Engineering Sub-City Geos for a Hyper-Local Marketplace with Uber. [viewed 10 October 2023]. Available from: <https://www.youtube.com/watch?v=wDuKeUkNLkQ>.
- Vasilev, J., M. Kehayova-Stoycheva, B. Serbezova.** (2022). A GIS-Based Approach in Support of Monitoring Sustainable Urban Consumption Variables. Education, Research and Business Technologies. Proceedings of 20th International Conference on Informatics in Economy (IE, 2021), Smart Innovation, Systems and Technologies (SIST). Springer, Vol. 276, 237–247. DOI: 10.1007/978-981-16-8866-9_20.
- Xia, F. F.** (2022). GIS Software Product Development Challenges in the Era of Cloud Computing. In *New Thinking in GIScience*, 129–142. DOI: 10.1007/978-981-19-3816-0_15.

EXPLORING THE CAPABILITIES OF THE H3 GEOCODE SYSTEM DEVELOPED BY UBER TECHNOLOGIES

Abstract: *The research focuses on a practical geocoding system with near-equal area tessellation that divides the spherical earth surface using pentagon and hexagon shapes. The resulting almost equal-area tessellation provides intriguing possibilities for quickly determining equidistant distances and visualizing the so-called “heat maps”. The latter allows for the easy analysis of dynamic and rapidly changing data in terms of the distribution of measured values per unit area. The most important features of the H3 system are highlighted in this regard. This system is useful in geographic information systems related to transportation and flow management. The latter is supported by numerous program libraries that are distributed under an open-source license. The research can serve as a basis for choosing an appropriate geocoding system, in the need for analysis and visualization of dynamic, rapidly changing data associated with geographic location.*

Keywords: *H3, Uber Technologies, geocoding, program libraries, equal area tessellation*

Assoc. Prof. DSc Pavel Petrov
University of Economics – Varna
E-mail: petrov@ue-varna.bg