Project Proposal
for GT Rent - GT student rental in Atlanta

The incredibles: Cheng Chen, Qi Ge, Jiahao Luo, Kai Lu, Shuo Huang

1.Objective(Heilmeier question #1)

Along with increasing student population in Georgia Tech, students who prefer living off campus encounter accommodation problems. Rents, distances to school, living convenience, and neighborhood crime rates are main aspects concerning them.

Songnian's book, reviewing development of web-based GIS and mapping services and application[1], inspires us to implement different sources of data and services with Web-GIS. Since many tools and sources are available, integrating services to functionalize the non-distributed GIS is achievable.

Our objective is to integrate data from Zillow, Google, Yelp and Atlanta local crime database into a web application platform, help the GT students find better living places.

According to Zhou's study of commuting in the University of Los Angeles[2], the most prevalent transit tools are car (41.2%), public transportation (30.9%) and walking (24.8%). Since transportation in Atlanta is similar to LA's, we collect distances between school and properties in three aspects: driving, public transportation, and walking.

According to Phibbs' research, indicating that straight line distance is a reasonable proxy for travel time, especially those less than 15 miles[3], we restrain data collection range within area surrounded by I-285 highway.

Since accessibility to retailing services are quite important for evaluating properties' value[4], we search each property's neighborhood for these services and combine them with crime rates and distances to campus to build a ranking system.

2.Progress and limitation(Heilmeier question #2)

Currently progress:

1) Build a clear and reachable plan, along with goals of the project;
2) Collect and clean target estate locations and their vicinity details by Google Maps API;
3) Fetch related property information by Zillow API;
4) Gather crime informations last five years in Atlanta;
5) Collect rating and categorical informations of surrounding services by Yelp API.

However, there are drawbacks in data precision. The limitation of Zillow API restrained us from fetching estate locations massively, so we searched for properties with Google Maps API. But the datasets collected are smaller than those displayed on Zillow for the same area. To overcome this defect, we adopted ideas from Crawling hidden objects, addressing the problem of crawling all objects efficiently from a location based service through public kNN web search interface[5]. We aim to realize 2D crawling algorithm and combine multiple places types together, a promotion to simply crawling LBS.

3.Approaches(Heilmeier question #3)

1)Yelp

Data from Yelp can evaluate whether there are enough retailing services near rental properties.Longitude and latitude of the houses and radiuses, one for walk and another for driving, are keywords to search.

The paper by Pai Peiyu and David C. Arnott introduces relationships between users and social networks sites[6], helping us understand our potential users. Amy Hicks's research examines people's motivations for using Yelp and their differences, including methods and needs[7], giving us data correlated with students. Weijia Dai's paper introduces cleaning data method for those containing restaurants on Yelp[8]. Mukherjee's paper[9] focuses on circumstances of possible fake reviews and how Yelp reacts to them, helping us evaluate the accuracy of Yelp reviews.

2)Google Maps

Crawling addresses' concepts are based on paper [10], from which we utilize similar idea about spatial range queries. Hybrid cumulated kNN searches searched for places, covering areas surrounded by I-285.

3)Crime rate

Crime data are downloaded directly. After learning basic crime data mapping concepts (repeat victimisation) in Hirschfield's book[11], we use Google Maps Geocoding API, converting addresses into geographic coordinates and making calculating crime rate in a specific area easier. Ultimately, we focus on relationships between crime rates and locations.

Devin's papers demonstrates the inverse relationship between crime rates and property values: "the estimated elasticities of property values with respect to crime range from − 0.15 to − 0.35"[12]. Leigh's research provides more detailed price gradient of distance from criminal offenses: a crime event has a -4% price reduction on houses within 0.1 miles, yet has no influence on houses farther away[13]. Thus, it is reasonable to rank the final rental result with the elastic numbers and distance gradient considering their linear relationship.

4)Zillow

The Zillow API method locates an estate based on specific address. It returns rental fee, lot size, year built etc. By Gelman's paper[14], which studies the difference between data qualities based on user-contributed information, Zillow improves the completeness and integrity instead of accuracy. Because of large and complex geographic data, we apply data mining techniques in Miller's book[15], including classification, clustering, association, trends and regression analysis.

4)Ranking

The basic model for decision making is Location Based Services[16]. In this paper, LBS runs the entire gamut from mapping services (Google Maps) to restaurants (Yelp) and real-estate (Redfin) with 2-dimensional kNN interface. We build the HDBSCAN-1D ranking model, with additional crime data.

4.Potential users(Heilmeier question #4)

In the short run, potential user for our project will mainly be students in GT. In the long run, the potential user could be people seeking apartments in general.

5.Comparison(Heilmeier question #5)

Details about rental properties' locations, rents, neighborhoods and residents' reviews are usually scattered throughout several sources: Google Maps, Yelp, and rental property official websites, making the normal searching process tedious and difficult. However, compared to Zillow and Trulia, which are unnecessarily complicated, our application combines all necessary features and information in a more concise and dedicated way, providing students with pleasant searching experience.

6.Risks(Heilmeier question #6)

Underestimating time, lacking experience, and information precision may cast shadow on the final result. It is challenging to implement web-based application creation in one month. Also, though we tried our best to refine the data, they may still contain dirty data points, causing inaccuracy.

7.Payoffs(Heilmeier question #6)

This cooperating and collaborating project benefits both us and users. We acquired new skills through cooperative learning such as future planning, mass data processing, and web application creation.

Though our project solely focuses on GT students, we can extend our work to any colleges in U.S, benefiting more college students in the future.

8.Cost(Heilmeier question #7)

Direct costs:

Device cost: personal desktop and notebook;

Indirect costs:

Possible cost: crawling more data than the bar each day;

Time cost:

On average 12 hours a week beyond class time.

2 hours Group meeting each week.

9.Project plan(Heilmeier question #8 & #9)

Mar.9 -Mar.18:

Discover correlation coefficient between different affecting factors and, ultimately, finding a formula and producing the best rental ranking.

Factors include position (by Ge and Lu), price (Luo), crime rate (Huang), restaurants and entertainments (Chen).

Mar.19 - Apr. 19

Build a webapp which enables users to reweight the ranking factors.

1) Choose platform, domain name and webhosting for our website.
2) Discuss type of web design.
3) Create web app and add functions.(by Lu, Luo and Ge)
4) Connect database to our web app.(by Chen and Huang)

Reference:

1. Hardie, Andrew. "The development and present state of Web-GIS." *Cartography 27.2 (1998): 11-26.*
https://www.crcpress.com/Advances-in-Web-based-GIS-Mapping-Services-and-Applications/Li-Dragicevic-Veenendaal/9780415804837

2. Zhou, J., 2012. Sustainable commute in a car-dominant city: Factors affecting alternative mode choices among university students. *Transportation research part A: policy and practice, 46(7), pp.1013-1029*
http://www.sciencedirect.com/science/article/pii/S0965856412000651

3. Phibbs, C.S. and Luft, H.S., 1995. Correlation of travel time on roads versus straight line distance. *Medical Care Research and Review, 52(4), pp.532-542.*
http://mcr.sagepub.com/content/52/4/532.short

4. Song, Y. and Sohn, J., 2007. Valuing spatial accessibility to retailing: A case study of the single family housing market in Hillsboro, Oregon. *Journal of Retailing and Consumer Services, 14(4), pp.279-288.*
http://www.sciencedirect.com/science/article/pii/S0969698906000488

5. Yan, H., Gong, Z., Zhang, N., Huang, T., Zhong, H. and Wei, J., Crawling Hidden Objects with kNN Queries.
http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7335622&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D7335622

6. Pai, Peiyu, and David C. Arnott. "User adoption of social networking sites: Eliciting uses and gratifications through a means–end ap*proach." Computers in Human Behavior 29.3 (2013): 1039-1053.*
http://www.sciencedirect.com/science/article/pii/S074756321200177X

7. Hicks, Amy, et al. "Why people use Yelp. com: An exploration of uses and gratifications." *Computers in Human Behavior 28.6 (2012): 2274-2279.*
http://www.sciencedirect.com/science/article/pii/S0747563212001951

8. Dai, Weijia, et al. Optimal aggregation of consumer ratings: an application to yelp. com. No. w18567. *National Bureau of Economic Research, 2012.*
http://www.nber.org/papers/w18567

9. Mukherjee, Arjun, et al. "What yelp fake review filter might be doing?."*ICWSM. 2013*.
http://www2.cs.uh.edu/~arjun/papers/ICWSM-Spam_final_camera-submit.pdf

10. Bae, W.D., Alkobaisi, S., Kim, S.H., Narayanappa, S. and Shahabi, C., 2009. Web data retrieval: solving spatial range queries using k-nearest neighbor searches. *Geoinformatica, 13(4), pp.483-514.*
http://link.springer.com/article/10.1007/s10707-008-0055-2

11. Mapping and Analysing Crime Data: Lessons from Research and Practice *(edited by Alex Hirschfield, Kate Bowers, April 26, 2001 by CRC Press, ISBN 9780748409228 - CAT# TF2275)*

12. Crime and property values: Evidence from the 1990s crime drop. *Devin G. Pope a, Jaren C. Pope.Regional Science and Urban Economics 42 (2012) 177–188*
http://www.sciencedirect.com/science/article/pii/S0166046211001037

13. Estimates of the Impact of Crime Risk on Property Values from Megan's Laws. *Leigh Linden and Jonah E. Rockoff. American Economic Review 2008, 98:3, 1103–1127*
https://www.aeaweb.org/articles.php?doi=10.1257/aer.98.3.1103

14. Gelman, Irit Askira, and Ningning Wu. "Combining structured and unstructured information sources for a study of data quality: a case study of Zillow. com." *System Sciences (HICSS), 2011 44th Hawaii International Conference on. IEEE, 2011.*
http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5718619&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5718619

15. Miller, Harvey J., and Jiawei Han, eds. *Geographic data mining and knowledge discovery. CRC Press, 2009.*
https://www.crcpress.com/Geographic-Data-Mining-and-Knowledge-Discovery-Second-Edition/Miller-Han/9781420073973

16. Rahman, M.F., Liu, W., Suhaim, S.B., Thirumuruganathan, S., Zhang, N. and Das, G., 2016. *HDBSCAN: Density based Clustering over Location Based Services. arXiv preprint arXiv:1602.03730.*
http://arxiv.org/pdf/1602.03730v2.pdf