The Millennium Project: Visualizing Future Impact of Technology in the World 2020 -2050

Client: Jerome Glenn, Co-founder (1996) and Director of 'The Millennium Project'

Team members:
Priyanka Jha Indiana University **prijha@.iu.edu**,
Julie Diby Indiana University **jdiby@indiana.edu**
Ian Bass Indiana University **ibass@indiana.edu**
Arthi Anand Indiana University **arthanan@indiana.edu**
Archana Singh Indiana University **arcsingh@iu.edu**

Fig. 1. The millenuim project interface

Abstract— While Technology is believed to have a relatively large impact on the future, technology also affects structural changes, especially on jobs and specifically on the unemployment rate. Today's questions will tend to resolve future problems as technology grows in greater competition with humans' skills and impacts future unemployment rates. In this project, we envisioned analyzing, in detail, some important socio- and politico-economic questions through visualizations and gaining useful insights.

Our client requested a visual representation of the Millennium Project on the Real Time Delphi (RTD) survey results regarding technology's impact on the future in the years 2020-2050. The visualizations would be a better way to garner insights into technology experts' visions and fears of the future. In this project, we explore participants' opinions on 1) The unemployment rate over time, 2) The geographical distribution of technology, 3) The occupational distribution of survey participants, and 4) Survey participants' concerns. We pre-processed and cleaned the data in Excel and sci2, then used Excel tool to further polish the data for establishing reflections and developing analysis using data exploration techniques. We then created a Dashboard for the visualization and exploration of several different perspectives on the survey using Tableau software.

Our client expressed great satisfaction, and we gained several insights such as: understanding of weak participation from Economists in the survey as they should be more concerned about the future and the significance technology plays in impacting jobs in the future; benefiting from better wealth distribution; understanding that the impact of some technologies like robotics on future unemployment is massive, but that of others like nanotechnology is limited. Ideally, the Millennium Project findings will be taken into account by stakeholders, entities, and mainly governments today to rectify the human challenges of tomorrow. A dynamic webpage was created for the client with an intent to help publish the visualizations and the insights on the Millennium Project website. Further explorations of these findings with additional data are possible.

Index Terms— Co-occurrence, Real Time Delphi data (RTD), word cloud, temporal analysis, topical analysis, geospatial analysis, and climate science literate.

INTRODUCTION

The Millennium Project (think tank) is a global foresight project that anticipates building a better future by improving humanity's prospects. The Millennium Project assumes that the world needs to think seriously about these challenges now, because it may take a generation or more to make the changes necessary to improve our future prospects. To address this challenge, they launched a Future Work/Technology 2050 study known as the Real-Time Delphi International Survey, where well-versed professionals in their fields gave their opinions about diverse technologies and their impact in projected-view time periods on jobs and unemployment.

Participants responded to ten questions as part of 'Future Work/Technology 2050 Real-Time Delphi Study'. We studied diverse opinions on technology impact and distributions with well-elaborated visualizations that give a clear view of how exactly these experts see the world 2020-2050 in regards to unemployment.

1. CLIENT'S REQUIREMENTS ANALYSIS

Our Client gave us freedom to analyze data as we wished but we discussed some directions and received guidelines and mainly focused our analysis on these questions:

1. Analyze unemployment change from 2020-2050, on survey respondents' level of expertise and age range.
2. Analyze the future distribution of all job-replacing technologies around the world based on participant's viewpoint in RTD survey.
3. Analyze the occupational self-identification of survey's participants.
4. Analyze the responses of participants on a series of questions, specifically common words and interpretation of participants' concerns.

2. RELATED WORK

Our client, Mr. Glenn- through his work on the Millennium Project- wants to raise awareness about the state of the future. Likewise, Niepold, Herring, and McConville also want to make the world more "climate science literate" [1]. They believe that climate research data can make the public more aware of climate changes and push the public towards tackling the future problems on CO^2 emissions in the world. Their study provided some inspiration on our animated geospatial visualization when addressing the Millennium Project data.

Additionally, Tiffany Holmes- Associate Professor at the School of Art Institute of Chicago- presents the artwork of 7000 oaks and counting. "The artwork consists of a lobby-sited kiosk and website that display an animation of dynamic energy loads at the National Center for Supercomputing Applications [2] (NCSA)" (Holmes, pg. 1, 2007). The aim of her project is to increase awareness among building residents to realize the carbon footprint they are leaving and possibly change their behavior (Holmes, 2007). The aim and methods employed in her project definitely resonate with the aim of the Millennium Project to increase awareness of problematic issues in the world.

3. DATA COLLECTION AND PREPARATION

The data came to us unrefined and we used data mining techniques to restore data logic so records /observations were placed in the right columns/fields. The data came to us in a comma-separated value format. However, a few of the columns had multiple survey responses which are also separated by commas and surrounded by brackets. Thus, our aim was to separate out any data that was in between brackets i.e. "[...]" by using Excel. We utilized the "text to columns" function with an open bracket ("[") as the data separator and later a close bracket ("]"). Then, within the new columns that were created (that had previously been inside of the brackets), we replaced all commas with pipe separators ("|"). We then aggregated the data back together within CSV format and were then able to parse the data in its intended format. Finally, we removed any remaining brackets and quotations ([,",]) and saved the file**.** More data transformations were done in Excel, including the average aggregation of the technology column and used multiple fields for the visualizations (same action using Tableau average function).

The dataset used for this project was imported from the Millennium Project website called themp.org. The data obtained is a CSV file with 4367 rows and 83 columns containing the Real-Time Delphi international survey data. It contains the user demographic information and the survey questions along with the user response captured.

The major entities are User, Question, and Response. User has attributes such as id, location, age, occupation, and institution. Most questions have an attribute called sub-question, where survey respondents were able to supplement their responses with additional text to elaborate on their viewpoints. Responses are in a collection of columns (subset of columns J through CE) that are mapped to a particular question/sub-question pair. One data row captures the response of a question/sub-question for one user. Since one user answers (ideally) 10 questions, data for one user ranges over several rows.

Keys Processes and Algorithms

After data restoration and cleaning process, data was transformed to fit each data analysis and visualization's requirements accordingly:

-RTD Temporal Analysis: data was aggregated by year in Excel. Averages of percentage employment for the years 2020, 2030, 2040, and 2050 would be a good reference but would not be a true representation of experts' viewpoint. The team decided to capture this data by also grouping on the various age groups, level of expertise, and occupational self-identification of futurist vs non-futurist. This would give an idea of the outlook of these age groups on employment changes from 2020 through 2050. The cleaned dataset was imported into Tableau. The unemployment averages values were grouped by age and experience using the same color scheme for the years and presented on a visualization dashboard to give a holistic view.

-RTD geospatial analysis: this visualization was suggested by the client to capture the anticipated spread of impactful technologies across the world. Initially, the geospatial visualization was focused on a world's view of nanotechnology, comparing two regions of the world (South America and Europe) and among profession/expertise level. The study was later shifted towards Robotics technology on a world's view approach and analyzes the view of this future technology by profession and expertise level. Based on the latest comments from the stakeholders, the team decided to show a dynamic visualization of the anticipated impact of all technologies with a dropdown list, thus allowing the client to select any technology. Data was transformed to show average perceived technological impact by country of the survey participant on a choropleth map in Tableau, using a two-color gradient color scheme to show average technology.

-RTD topical analysis: we initially planned on a Word Co-Occurrence Network analysis, showing the relationship/link between futurists and non-futurists. This was later refined for a more meaningful approach to visualize survey respondents' self-identified occupations and which ones occurred together most often. Excel was used to consolidate survey responses into each responder rather than all responses to all questions. We used the "Extract Word Co-Occurrence Network" algorithm in Sci2, used a v-lookup to attach the node numbers to the node names (listed Occupations), then mirrored the original NWB output to make the data bi-directional, and finally visualized in Tableau. This version provides more quantifiable information, as many survey participants listed more than one occupation.

-Word cloud: an online word cloud generator (wordclouds.com) was used to create the word clouds. The generator takes the input file in text format and generates word clouds on different shapes and background colors. While the algorithm used isn't given, the generator outputs the words used and each word's frequency. This gives the user the option to remove any word. The two parameters used were: **Direction**: Horizontal, and **Shape**: Cloud.

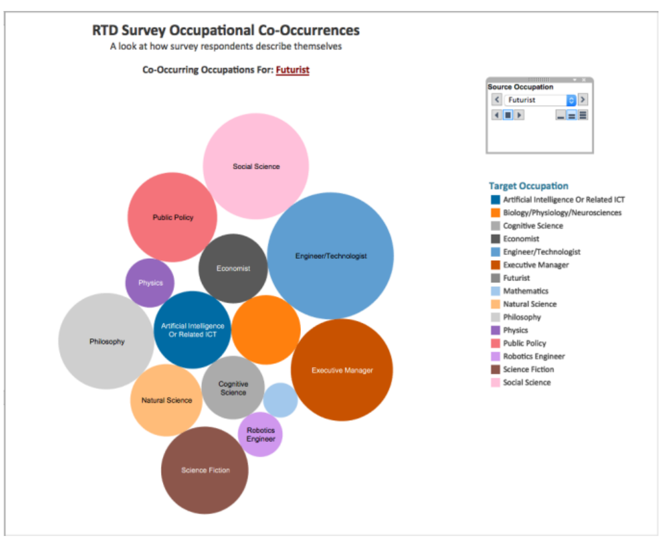
4. ANALYSIS AND VISUALIZATION

The outputs generated from the analysis were 5 different visualizations combined for final visualization.

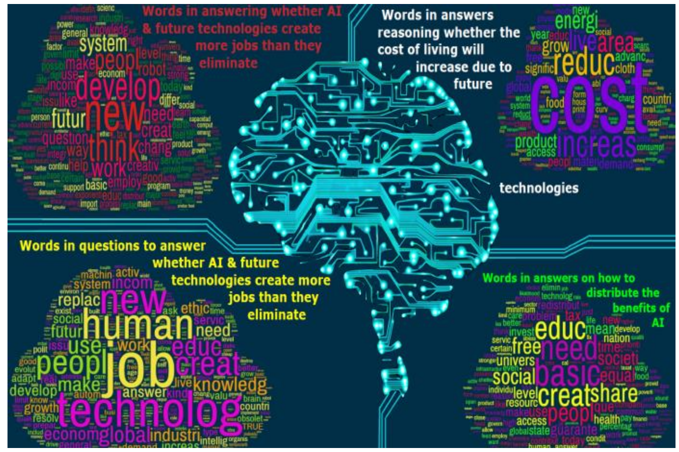
- *A Temporal analysis of future unemployment:* this dashboard captures the outlook of experts' viewpoint on unemployment changes from 2020-2050 by age group as well as their stated level of expertise. Additionally, the client requested a Futurist vs. Non-Futurists outlook on Unemployment, which was later implemented in the visualization.

- A *Geospatial visualization*: presents the spread by country on a world map, the technology expected to replace jobs in the future.

- *A Topical Visualization*: mapping occupational distribution of participant (As participants describe themselves as having other functions as well.)



- *A Word Cloud*: analysis of recurring words from answers to questions 3.1, 3.2, 5.1 and 8.2 (Please find respective questions in Technical Appendix section)



5. KEY INSIGHTS GAINED FROM VISUALIZATIONS

5.1 RTD Temporal Visualization Insights –

Survey respondents in all age groups expect to see a pronounced increase in unemployment from 2020 to 2050. Participants of low, medium, and high experience also foresee an increase in unemployment from 2020 to 2050. What is particularly interesting is that higher experience participants foresee a higher increase in employment as compared to medium and low experience participants. Also, Futurists predict a greater future unemployment rate than the Non-Futurists.

5.2 RTD Geospatial Visualization Insights –

Across regions, all participants agree on the Robotics revolution in the coming years, according to the responders. Idem for analysis across different professions and different expertise levels regarding Robotics technology, same findings. The survey results indicate that Robotics will be increasingly and widely used in the future. Robotics is the most popular and promising technology of the future with the highest average impact of 7.51 and a leading factor in creating unemployment. Our results corroborates the client's findings.

5.3 RTD Topical Visualizations Insights –

Many survey participants cited more than one occupation. For example, the below occupation visualization screenshot shows how respondents describing themselves as 'Futurists' also share other occupations and what the distributions are. Observations: weaker than expected participation of Economists, as one would expect from a survey on future unemployment. As expected, the occupation of 'Futurist' dominated the survey responses and was most commonly associated with 'Engineer/Technologist' and, perhaps, somewhat surprisingly, 'Social Science'.

5.4 RTD Word Cloud Insights –

The word cloud created was related to questions 3.1 and 3.2, which were essentially questions and answers of experts across the world to understand whether AI and other future technologies can create more future jobs than they eliminate. Experts used words such as "human" "AI", "new", "technology", "Jobs", "Social", and "National" more often. Overall, experts across the world were open to accept these future technologies in the belief that jobs will be created and people will be progressive.

There are two additional word clouds, the first for question 5.1, which answers how the cost of living is impacted by the future technologies. There were conflicting views regarding the impact. Some believed that AI would reduce the cost of living while others disagreed. Words such as "demand", "people", "food", and "popul"(stem for words such as population, populated, etc.) have been used. The main insight derived from the visualization is that people are considering issues such as demand for materials, consumption, food, and wealth created by technologies as important in deciding the impact on cost living.

From the second additional word cloud on question 8.2, it can be understood that responders perceive that the benefits of AI must be distributed equally to all classes in order to fulfill basic needs, such as education, food, and shelter. Second, words such as “redistribute”, ”share”, and ”equal”, reinstate that the distribution would initially cause a gap between the different levels in society and thus the wealth must be redistributed to all. The next levels benefits must be shared such that there is good infrastructure, access to resources, improved working conditions, pay, and health.

Insights on Further Exploration

Some of the key challenges are the direction to take from this information, and there are many avenues to further explore this data. The survey could have been expanded on indexes of development or measurable values for each continent/region, which could have better supported the participant’s point of view. The complexity involved was mostly in coming up with a creative way to get the data to show something meaningful. There is an opportunity for more extensive work, insights, and visualizations to be created with the available resources. Under our client’s guidelines and after final submission to the client, different, compounded, or further refined visualizations could be explored.

6. VALIDATION AND CONCLUSION

Multiple internal team meetings and meetings with the clients were arranged to develop a consistent understanding of the data, requirements, and to showcase our findings in visually appealing ways. Feedback from our client, Jerome Glenn, and Prof. Börner from Indiana University was incorporated on a regular base, and all prototypes have been officially approved by our clients. Access to the current versions of our visualizations was provided to the client to run through the visualization.

Mr. Jerome Glenn expressed great satisfaction about our multiple data analyses based on excel, the Sci2 tool, and visualizations based on Tableau. We gained several insights during this process that are worth aforementioned. Client attested that there are so many data manipulations and visualizations that can be made. Client recognizes that there is an infinite number of ways to present the whole project. The client was delighted with visualizations.

ACKNOWLEDGMENTS

We would like to deeply thank our client, Mr .Jerome Glenn, Director of ‘The Millennium Project’; Professor Katy Börner, Michael Ginda, and Ashish Shendure of Indiana University for their prompt feedback and insights on preparing the project. Thanks as well to the peer reviewers for their participation and constructive comments; the team, Priyanka Jha, Julie Diby, Ian Bass, Arthi Arnand, and Archana Singh for their dynamic contribution and for working closely and in symbiosis.

REFERENCES

[1] Niepold, Frank, David Herring, and David McConville. "The role of narrative and geospatial visualization in fostering climate literate citizens." *Physical Geography* 29.6 (2008): 529-544.

[2] Holmes, Tiffany Grace. "Eco-visualization: combining art and technology to reduce energy consumption." *Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition*. ACM, 2007

[3] http://www.breathingearth.net

[4] https://themp.org/

[5] http://millennium-project.org/

[6]http://wiki.cns.iu.edu/display/SCI2TUTORIAL/Science+of+Science+(Sci2)+Tool+Manual

[7] https://www.jasondavies.com/wordcloud/about/

TECHNICAL APPENDIX

A. **Co-occurrence network:**
The topic similarity of basic and aggregate units of science can be calculated via an analysis of the co-occurrence of words in associated texts. Units that share more words in common are assumed to have higher topical overlap and are connected via linkages and/or placed in closer proximity. Word co-occurrence networks are weighted and undirected. (Source: http://wiki.cns.iu.edu)

B. **Real time Delphi data / RTD:**
The Real-Time Delphi is a relatively new and efficient method for collecting and synthesizing expert opinions. The original Delphi technique was developed by the RAND Corporation in the late 1950s. Although it has produced many valuable insights for many around the world, it requires multiple rounds of questionnaires that can take months to complete. The big advantage of the RTD is that it is a “roundless” Delphi. There is no need for an explicit second round. The respondents participate by filling out an online questionnaire, and the results—both numerical and qualitative—are updated as responses are recorded in “real time.” Respondents can—and are encouraged to—revisit the questionnaire as many times as they want. Each time, they are shown their own responses as well as the updated answers of the others, and they can revise and change their own inputs based on this feedback. (Source: http://www.millennium-project.org)

C. **Word Cloud:**
A word cloud is a popular visualization of words typically associated with Internet keywords and text data. They are most commonly used to highlight popular or trending terms based on frequency of use and prominence. A word cloud is a beautiful, informative image that communicates much in a single glance. (Source: http://www.peachbelt.com)

D. **Temporal Analysis:**
Temporal analysis aims to identify the nature of phenomena represented by a sequence of observations such as patterns, trends, seasonality, outliers, and bursts of activity.
A time series is a sequence of events or observations that are ordered in time. Time-series data can be continuous (i.e., there is an observation at every instant of time) or discrete (i.e., observations exist for regularly or irregularly spaced intervals). Temporal aggregations – over journal volumes, years, or decades – are common.
Frequently, some form of filtering is applied to reduce noise and make patterns more salient. Smoothing (i.e., averaging using a smoothing window of a certain width) and curve approximation might be applied. The number of scholarly records is often plotted to get a first idea of the temporal distribution of a dataset. It might be shown in total values or as a percentage of those. One may find out how long a scholarly entity was active; how old it was at a certain point; what growth, latency to peak, or decay rate it has; what correlations with other time series exist; or what trends are observable. Data models such as the least squares model – available in most statistical software packages – are applied to best fit a selected function to a data set and to determine if the trend is significant. Kleinburg's burst detection algorithm is commonly applied to identify words that have experienced a sudden change in frequency of occurrence. (Source: http://wiki.cns.iu.edu)

E. **Topical Analysis:**
Topical analysis extracts the set of unique words or word profiles and their frequency from a text corpus. Stop words, such as 'the' and 'of' are removed. Stemming can be applied. Co-word analysis identifies the number of times two words are used in the title, keyword set, abstract and/or full text of a paper. The space of co-occurring words can be

mapped providing a unique view of the topic coverage of a dataset. Similarly, units of science can be grouped according to the number of words they have in common.
Salton's term frequency inverse document frequency (TFIDF) is a statistical measure used to evaluate the importance of a word in a corpus. The importance increases proportionally to the number of times a word appears in the paper but is offset by the frequency of the word in the corpus.
Dimensionality reduction techniques are commonly used to project high-dimensional information spaces (i.e., the matrix of all unique papers multiplied by their unique terms, in a low, typically two-dimensional space). (Source: http://wiki.cns.iu.edu)

F. **Geospatial Analysis:**
Geospatial analysis has a long history in geography and cartography. Geospatial analysis aims to answer the question of where something happens and what impact that something has on neighboring areas.
Geospatial analysis requires spatial attribute values or geolocations for authors and their papers, extracted from affiliation data or spatial positions of nodes, generated from layout algorithms. Geospatial data can be continuous (i.e., each record has a specific position) or discrete (i.e., each set of keywords has a position or area-shape file – e.g., number of papers per country). Spatial aggregations (e.g., merging via ZIP codes, counties, states, countries, and continents) are common.
Cartographic generalization refers to the process of abstraction such as (1) graphic generalization: the simplification, enlargement, displacement, merging, or selection of entities without enhancing their symbology; and (2) conceptual symbolization: the merging, selection, and symbolization of entities, including enhancement – such as representing high-density areas with a new (city) symbol.
Geometric generalization aims to solve the conflict between the number of visualized features, the size of symbols, and the size of the display surface. Cartographers dealt with this conflict intuitively in part until researchers like Friedrich Töpfer attempted to solve them with quantifiable expressions. (Source: http://wiki.cns.iu.edu)

G. **Climate Science Literate**
Climate Science Literate is one who has an understanding of your influence on climate and climate's influence on you and society. A climate-literate person, and understands the essential principles of Earth's climate system, • knows how to assess scientifically credible information about climate, communicates about climate and climate change in a meaningful way, and is able to make informed and responsible decisions with regard to actions that may affect climate.

H. **Question 3:** What questions have to be resolved to answer whether AI and other future technologies create more jobs than they eliminate?

I. **Question 3.1:** What questions have to be resolved?

J. **Question 3.2:** What are your thoughts about answers to the questions you suggested?

K. **Question 5.1:** What is likely to happen by 2050?

L. **Question 8.2:** If the future AI/Robot economy creates the abundant wealth many expect by 2050, how should it be distributed? Please explain briefly how this might be accomplished.

M. **Question 1:** If socio-political-economic systems stay the same around the world, and if technological acceleration, integration, and globalization continue, what percent of the world do you estimate could be unemployed—as we understand being employed today—during each of the following years: 2020; 2030; 2040; 2050?

N. **Question 2.1**: Factors replacing more jobs/work by 2050, preventing mass unemployment rated on a scale from 10 as Primary Cause to 0 No Impact at All.