TreeeX: Exploring the Diversity of Tree Species

S. Jänicke

Image and Signal Processing Group, Institute of Computer Science, Leipzig University, Leipzig, Germany

Abstract

The GlobalTreeSearch database provides a mapping between 60,065 tree species known to science and the countries where these trees grow. TreeeX is a visual exploration system that supports multifaceted analyses of the GlobalTreeSearch data. Focusing on the entire earth or a group of countries, investigating research questions on biodiversity are visually supported by interactive choropleth maps that color countries according to frequency, diversity, or uniqueness of prevalent tree species. Focusing on a single country, similarities and differences to other countries can be analyzed in detail. Several examples outline the system’s capability of delivering insights concerning the geographical diversity of tree species.

1. Introduction

The importance of trees for the ecological system has been discussed in a multitude of publications. As trees are at the base of the trophic pyramid [KS84], they frame the habitat of other species that depend on prevalent tree species. Taking the human as example [FPZ14], we need trees as resources for food, for building material, for medicine, and we often choose woodlands for recreational purposes. Also, the significant role of trees for stabilizing the Earth’s climate system, thus, the necessity of preserving forests and woodlands is a well-known fact [PvGST15].

In order to assess, monitor, and manage tree species diversity on a global, regional, and/or national level, researchers of Botanic Gardens Conservation International (BGCI) combined over 500 published sources on regional, taxonomic, and country-specific tree growth [BROS17] and made the resultant geographical mapping of tree species available in the form of the GlobalTreeSearch online database [BGC18]. Users of the database are provided with a textual search interface that allows for searching tree species and/or tree genera for a given country. The search result is a textual extract of the database listing family, taxon name and corresponding author of the tree species that match the given query; Figure 2 shows an example output after searching for hawthorn species in Germany. Though lists for individual queries can be downloaded in a tabular format, a comparative analysis of the geographical diversity of tree growth is not supported.
TreeeX was designed in order to provide multifaceted visual access to the GlobalTreeSearch data. As the geographical mapping is given on the country-level, TreeeX supports analyses on the biodiversity of tree species with interactive choropleth maps that illustrate different scenarios. Next to comparing the distribution of different tree species, the system allows for exploring aspects on species richness and uniqueness. In addition, similarities and differences of a selected country compared to other countries can be analyzed.

2. Related Work

Geovisualizations as means of illustrating aspects on biodiversity have been frequently used; exhaustive overviews of related techniques from thematic cartography and geovisualization are provided by Slocum et al. [SMKH09] and by Andrienko et al. [AA05].

Usually, thematic maps are based on a qualitative data set for a small geographical area, be it a national park or a whole country. Debinski et al. [DKJ99] use a geographical information system (GIS) to categorize habitats, and to determine relationships between remotely sensed habitat categorizations and species distribution patterns in the Greater Yellowstone Ecosystem. Thematic maps have also been used for a qualitative study of vegetation in Madagascar to draw implications and recommendations for the conservation of biodiversity [DPM98]. Similarly, Madden [Mad04] uses various heat maps to assess vegetation patterns in Great Smoky Mountains National Park. Setturu et al. [SR16] use categorical maps to visualize landscape dynamics in National Parks of Central Western Ghats. Categorial pixel-based dot maps, where each tree species or genera in the study receives a certain color, can be used to visualize detailed information on the regional distribution of tree species [RFN*08, HHP*16]. Predominance tag maps can be used to overlay such maps with textual information on locally predominant tree species [RCSJ18].

Geovisualizations are further central to analyzing time-dependent tree cover changes. Carnaval et al. [CM08] use heat maps to model the spatial range of the Brazilian Atlantic forest under three climatic scenarios (current climate, 6000 and 21,000 years ago) in order to predict patterns of current biodiversity. Based on analysis of nearly 30,000 Landsat images, Kim et al. [KSN*14] illustrate worldwide forest-cover change from 1990 to 2000 with heat maps. Heat maps are also used by Hansen et al. [HPM*13] to visualize tree cover, forest loss, and forest gain, and by Allen et al. [ABM15] for the analysis of worldwide locations of substantial tree mortality and forest die-off from hotter drought in the Anthropocene.

Linked views that interlink geospatial and temporal information are also valuable for exploring biodiversity data. Jänicke et al. [JS14] use GeoTemCo [JHSS12, JHS13] for the geospatial-temporal mapping and biodiversity analyses of the bentic invertebrate fauna. CommonGIS, a similar system, is used to combine forest ecosystem models with exploratory data visualization for the analysis of long-term simulation results [CKM*05].

Image processing techniques are used for generating information about the diversity of trees in forests. While Clark et al. [CRC05] use high spectral and spatial resolution imagery for the automated species-level classification of individual tree crowns in a tropical rain forest, Simons et al. [SHTA14] propose a point-based rendering method to support the remote sensing of forests. Musasabi [KKO15] aims at simulating the growth and changes of forests depending on the species of the tree and the land conditions.

3. TreeeX System

TreeeX is composed of an interactive choropleth map that is colored according to three different analysis modes: species analysis, global diversity, and country of interest. A screenshot of the system is given in Figure 3.

3.1. Species Analysis

In order to give a first impression of the most frequent tree species in the database, a tag cloud is shown with words scaled accordingly. Especially for novice users, the tag cloud aids as a starting point due to their intuitiveness, and their widespread usage to display summaries of textual data [VW08]. When hovering a tag, the number of countries where the chosen tree species grows is shown, and on mouse click, the corresponding countries are highlighted in a certain color; tree species can also be directly requested with a keyword search. The interface enables comparing the distributions of different tree species visually. When multiple species are prevalent in a country, the different colors are overlaid using a subtractive color mixing scheme. As the number of colors that can be easily distinguished by humans is limited by 12 [War04], though extendable, comparing a maximum of three tree species is suggested, which results in a color map consisting of seven distinctive colors in the CMYK color model as illustrated in Figure 4.

![Figure 4: CYMK color mixing for the species A, B and C](Image)

© 2018 The Author(s)
3.2. Country of Interest

Focusing on a single country $c_i$, each country $c_j$ of the choropleth map is colored in red with a saturation level according to the similarity of tree species $T(c_i)$ and $T(c_j)$ of $c_i$ and $c_j$. Two coloring modes are possible. The Jaccard index defined as

$$J(c_i, c_j) = \frac{|T(c_i) \cap T(c_j)|}{|T(c_i) \cup T(c_j)|}$$

considers the cardinalities of both sets of tree species, thus, providing a general view on the similarity of the biodiversities of $c_i$ and $c_j$. On the other hand, when the user likes to see the countries sharing most of $c_i$’s tree species, thus, providing similar habitat conditions, the cardinality of $T(c_j)$ is disregarded and only the overlap size compared to the richness of tree species of $c_i$ is considered as

$$I(c_i, c_j) = \frac{|T(c_i) \cap T(c_j)|}{|T(c_i)|}.$$

For exploration purposes, each country is attached with a bar chart reflecting these proportions. The center bar colored green illustrates the amount of shared species, the orange bar on the left the amount of tree species only prevalent in $c_i$, and the purple bar on the right the amount of tree species only prevalent in $c_j$. When hovering the bar chart, a popup window is shown that provides several information, and it can be expanded for a more detailed investigation of individual tree species. An example focusing on Germany is shown in Figure 1(b).

3.3. Global Analysis

Various research questions on the biodiversity of tree growth can be investigated in this mode for a given set of countries. The choropleth map can be colored according to three different aspects of global diversity:

**Frequency.** Countries are colored in red with a saturation level according to the total of different tree species divided by the total of the country with most tree species. Currently, the database holds Brazil with 8,982 tree species as country with the highest biodiversity.

**Uniqueness.** Countries are colored in red with a saturation level according to the uniqueness of tree species, that is the number of tree species unique in a country divided by the country’s total of tree species. Currently, Madagascar is listed as the country with the most unique biodiversity as 3,087 out of 3,315 tree species only grow in Madagascar.

**Diversity.** The Country-of-Interest functionality gives a comparative view on the similarities of tree species of a chosen country compared to all other countries. In order to generate a global view on the changing biodiversity among the given set of countries, this mode takes the similarities of all country tuples at once into account. Therefore, an $n \times n$ distance matrix $D = [d_{ij}]$ for $n$ countries is calculated. $d_{ij}$ denotes the dissimilarity of tree species between the countries $c_i$ and $c_j$, which is defined as $d_{ij} = 1 - J(c_i, c_j)$. Then, a classical multidimensional scaling (MDS) algorithm [BG05] is executed that determines a position $p_i = (x_i, y_i, z_i)$ in the 3D space $S = [0,1]_x \times [0,1]_y \times [0,1]_z$ for each country $c_i$ while preserving distances between countries as well as possible. $S$ is then scaled to

\[ S' = \frac{S - \text{min}(S)}{\text{max}(S) - \text{min}(S)} \]
the RGB cube, so that the color for a country is determined by

\[ R_i = \frac{r_{\text{max}} - r_x}{r_{\text{max}}} \cdot x_i + \frac{r_x}{r_{\text{max}}} \cdot 255 \]

\[ G_i = \frac{r_{\text{max}} - r_y}{r_{\text{max}}} \cdot y_i + \frac{r_y}{r_{\text{max}}} \cdot 255 \]

\[ B_i = \frac{r_{\text{max}} - r_z}{r_{\text{max}}} \cdot z_i + \frac{r_z}{r_{\text{max}}} \cdot 255. \]

\( r_x \) defines the spanned range of all \( x \)-values, thus, determined as \( r_x = x_{\text{max}} - x_{\text{min}} \). \( r_y \) and \( r_z \) are defined equally, and \( r_{\text{max}} = \max \{ r_x, r_y, r_z \} \).

4. Working with TreeeX

In order to illustrate the value of the proposed system to investigate different research questions on the biodiversity of tree species, this section outlines typical workflows with TreeeX.

Species Analysis. Researchers already stated that around 58% of all tree species are single country endemics [BROS17]. When browsing the tree species for a certain country, those unique species are marked with a fingerprint icon. Taking Chile with a total of 158 species as an example, unique and common species are already visible in the popup window as illustrated in Figure 6a. The three pepper tree species prevalent in Chile are not unique, which deserves a more detailed investigation of the distribution areas with the species analysis functionality. The resultant choropleth map marks Chile as the only country where all three species grow (see Figure 6b). A similar example is given in Figure 1(a) that shows the distribution areas of the three hawthorn species that grow in Germany. While all three species are endemic central Europe, \textit{Crataegus monogyna} (common hawthorn) has the largest distribution area excluding the Baltic states.

Country of Interest. The role of Panama’s biodiversity has been the subject of interest due to the global importance of the Panama Canal [CRIn‘01]. Figure 7a illustrates a high similarity to the tree species prevalent in Costa Rica, and the farther away from Panama to the north or to the south the amount of overlap decreases. That Panama lies in the center of a continuum of a north-south biodiversity change is depicted in Figure 7b. Focusing on Germany 1(b)
shows that to the north, only few species are endemic that do not grow in Germany. On the other hand, the farther to the south, the more tree species occur that are not endemic in Germany.

Global Analysis. When comparing the diversities of tree species of all countries, thus, choosing the diversity view in the global analysis mode, yet regional similarities and geographical changes get visible. Looking at the entire earth (see Figure 5c), one can see several clusters of similarly colored countries sharing a certain amount of tree species. Expectedly, this is the case for countries belonging to the same continents, e.g., Europe and Africa. But also, countries of different continents receive similar shades illustrating a higher similarity of biodiversities. The diversity among these countries becomes visible on a smaller scale, thus, by assigning a larger color space. By first zooming to the Americas (see Figure 8a), and then, by zooming to South America, as shown in Figure 8b, five clusters of countries having similar biodiversities can be seen. Switching to the Country of Interest mode with a focus on Argentina confirms the diversity mapping with highest similarities to Paraguay to the Country of Interest mode with a focus on Argentina comes visible on a smaller scale, thus, by assigning a larger color similarity of biodiversities. The diversity among these countries becomes visible on a smaller scale, thus, by assigning a larger color space. By first zooming to the Americas (see Figure 8a), and then, by zooming to South America, as shown in Figure 8b, five clusters of countries having similar biodiversities can be seen. Switching to the Country of Interest mode with a focus on Argentina confirms the diversity mapping with highest similarities to Paraguay.

5. Conclusion

TreeeX is a system that supports the comparative geographical analysis of biodiversities on the country-level. For that purpose, TreeeX provides an interactive choropleth map that is colored according to the given research question. So, TreeeX enables a multifaceted visual access to a rich database fundamental for research on the biodiversity of tree species. Temporal information on biodiversity, which are not contained in the GlobalTreeSearch database, but pertinent for biodiversity research are currently not considered by TreeeX.

The TreeeX interface is designed the way that it can also be used for other kinds of data where geographical regions—countries, states, districts, etc.—have overlapping sets of entities. In order to foster future research, TreeeX is implemented as a user-configurable JavaScript library based on D3.js, and it will be made available as open-source soon.

References

Figure 8: Comparing biodiversities on different zoom levels


Figure 8: Comparing biodiversities on different zoom levels


