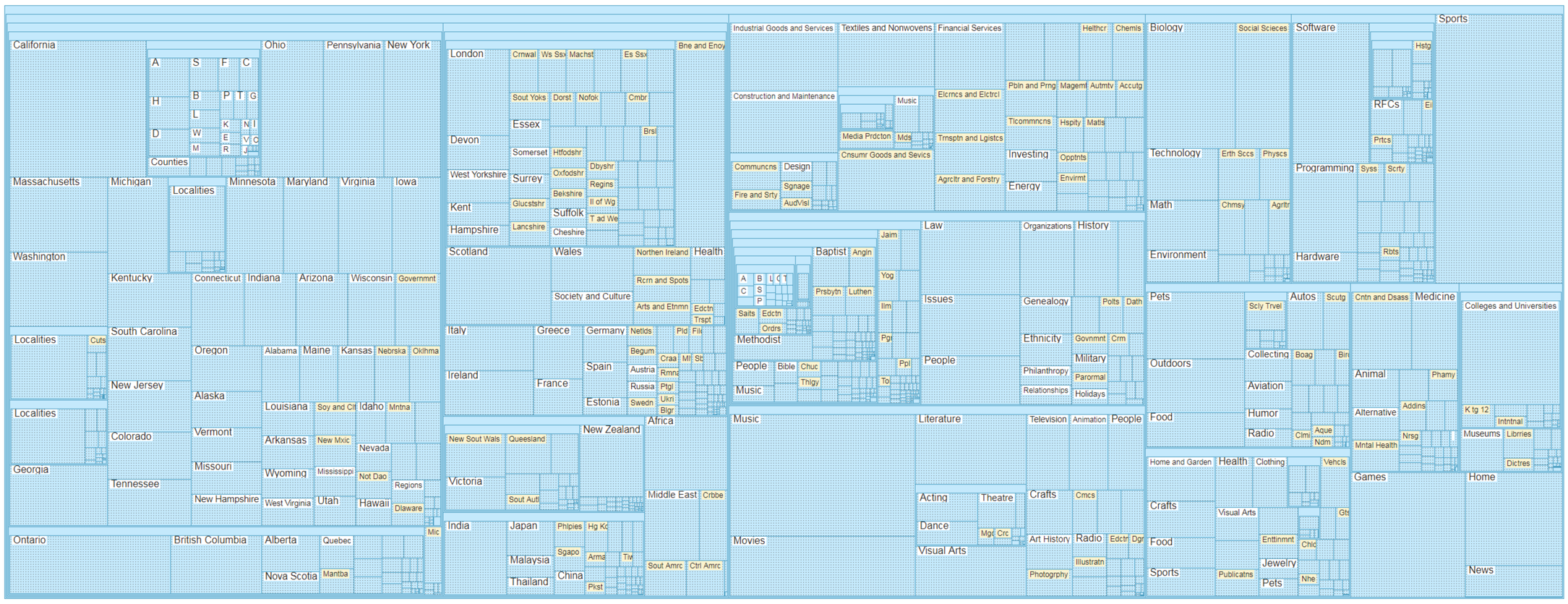
Abbreviating Text on Demand

Mariana Shimabukuro and Christopher Collins

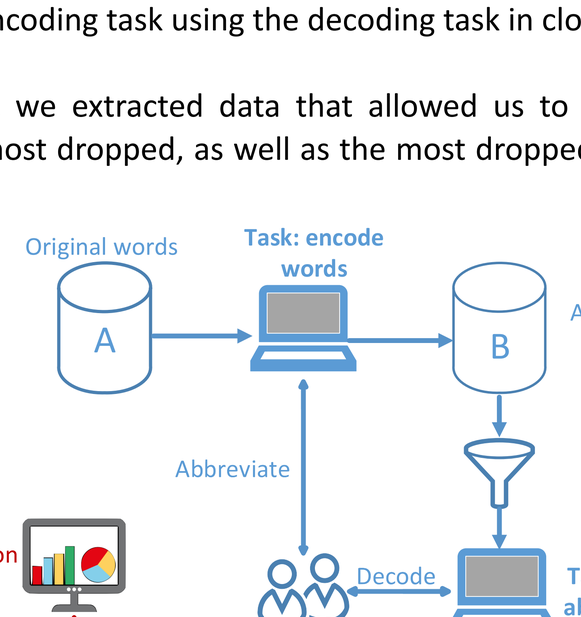
API available at: abbreviation.vialab.ca



A visualization of the DMOZ dataset. Each of the highlighted labels would not be displayed if they were not being abbreviated by our algorithm, which drops as many letters as needed to fit the text. It chooses the least important letter based on the character and its position within the word.

The Problem

Long text labels is a known challenge in information visualizations. Among the common techniques used to solve this problem are font size manipulation, wrapping sentences, dropping letters and text truncation.



Adaptive Crowdsourced Study

We ran the adaptive study on a crowdsourcing platform called Crowdflower with a total of 100 participants. By adaptive, we mean using the fast crowdsourcing recruitment and being able to evaluate the abbreviations created in the encoding task using the decoding task in close to real time.

From the study, we extracted data that allowed us to determine which letters are the most dropped, as well as the most dropped positions within a word.

Abbreviation on Demand Algorithm

Based on the study results we designed the "Abbreviation on Demand'' algorithm, which drops the least important letters of a word based on the study data, shortening labels while maintaining readability. The algorithm uses the probability of dropping letters based on their position within the word and the identity of the characters themselves.

In order to choose the letter to be dropped we calculate the following *score*:

$$score_{word[i]} = \begin{cases} monoDropProb(word[i]) * pPos(i) & \text{if } i = 0, \\ corrMx[word[i-1][word[i]] * pPos(i) & \text{if } i > 0. \end{cases}$$

where the *corrMX[word[i-1]][word[i]]* is the probability of dropping *word[i]* when it appears after the letter *word[i-1]* and *monoDropProb (word[i])* is the probability of a individual letter *word[i]* being dropped based on the study data. Considering that the correlation measure of a letter depends on the letter that came before, we cannot apply it to the first letter of a word.

Results

Abbreviations except TOP 1, 2 and 3 were created by dropping 40% of the letters from each word. Column "Original'' is the original word, followed by our algorithm and the other techniques. TOP 1, is the most accurate abbreviation from our study followed by TOP 2 and 3.

Original	**Abbreviation on Demand**	Drop Vowels	Truncation	Truncation keep end	TOP 1 decoded abbreviation	TOP 2 decoded abbreviation	TOP 3 decoded abbreviation	Original length	60% of length
academically	**acdmcly**	acdmcll	academi	academ.y	acad	academi	acdmcll	12	7
accelerating	**accelng**	acclrtn	acceler	accele.g	acclrtn	accelerat	acceler	12	7
acceleration	**acceltn**	acclrtn	acceler	accele.n	acc	accel	acceler	12	7
adventurers	**advturs**	advntrr	adventu	advent.s	advntrs	adventu	advnturers	11	7
assignments	**assgnms**	assgnmn	assignm	assign.s	assgnmnts	assign	assignments	11	7
atmospheric	**atmsphc**	atmsphr	atmosph	atmosp.c	atmsphr	atmsphrc	atmpheric	11	7
automotive	**autmtv**	autmtv	automo	autom.e	auto	automtv	autom	10	6
circumstance	**circmsc**	crcmstn	circums	circum.e	circums	crcstance	circmstnc	12	7
collisions	**colsns**	cllsns	collis	colli.s	collsns	collsn	collisi	10	6
colonization	**colnzan**	colnztn	coloniz	coloni.n	colonztn	colniztn	colnzation	12	7

Using our API with D3

```
this.select("node")
      .text(function (d){
         $.ajax({
            //using API to abbreviate d.label into 5 characters
            url: "https://abbreviation.vialab.ca/abbreviate?word="+d.label+"&length=5",
            async: false,
            success: function (data){ return data.abbr; }, //return abbreviation
            error: function (result){ return "";} //could not abbreviate
         });
      });
```

UNIVERSITY OF ONTARIO INSTITUTE OF TECHNOLOGY · vialab visualization for information analysis · NSERC CRSNG

• [marianaakemi.shimabukuro, christopher.collins]@uoit.ca • **API available at: abbreviation.vialab.ca**