# Tracing State Legislation to Its Source

Matt Burgess
University of Michigan

Julian Katz-Samuels,
University of Michigan

Lauren Haynes
University of Chicago

Eugenia Giraudy
University of California-
Berkeley

Joe Walsh
University of Chicago

**Extended Abstract**

## Problem

State governments have a central but understudied role in policy making. Each year, states spend 1.5 trillion dollars on programs and services for their citizens and pass 75 times more bills than Congress. However, state legislators often lack the time, staff, and expertise necessary to draft each bill. As a consequence, they often rely on legislation written by outside entities, such as interest groups or legislators from other states. Unfortunately, it is difficult for journalists, scholars, and other interested citizens to understand where those policy ideas come from. The average graduate student would need almost 900 days just to read the 70,000 bills introduced last year, let alone compare them. The time-consuming nature of this approach makes it incredibly hard for journalists or scholars to fully understand where state legislations ideas originate. As a result, citizens lack information about who is influencing state laws, affecting transparency and democratic accountability.

## Current approach

Existing approaches focus on manually reading legislation to find legislative text reuse across bills or from model legislation drafted by interest groups. As a result, journalists and researchers tend to limit the amount of bills by focusing either on federal bills or on a subset of topics (e.g. abortion) or interest groups (e.g. ALEC). Even Google searches can be slow: to find good matches the user has to find a part of the bill to search for and then look through the results, many of which are not legislation at all. Consequently, existing analysis tend to be limited and biased.

## Our Approach

To help journalists and researchers expand their knowledge on state politics we are building a tool to automatically detect cases of text reuse in state legislative bills. Previous efforts to build such a tool have failed because the required text comparisons are computationally costly. We simplify the problem by first limiting the number of comparisons to be made using a customized search engine built using Elasticsearch. Then we use the Smith-Waterman local alignment algorithm, similar to the one used for DNA sequence matching, to detect sequences of text that occur both in model legislation and state bills.

## Data

For our analysis, we use a number of distinct data sources. First, we use Sunlight Foundation's corpus of state legislation. The data include more than 500,000 bills and 200,000 resolutions for 50 states, ranging from 2007 to 2015. On average, each state introduced 10,524 bills, with an average length of 1205 words. Second, we have collected more than 1500 pieces of model legislation from groups across the political spectrum, including ALEC, the State Innovation Exchange, and the Uniform Law Exchange.

## Evaluation

To evaluate our system we created an evaluation set that includes 165 bills with legislative text reuse. To build this evaluation set we manually read through bills that were highlighted by journalists and experts as cases of text reuse. We then grouped bills into sets with matching text. Using this, we evaluated the ElasticSearch and local sequence alignment components of our system separately. First, for each document in the dataset, we queried ElasticSearch and calculated the recall of the results that met a certain pre-designated threshold. Second, on every pair of documents, we ran the Smith-Waterman algorithm using different parameter settings and inspected the precision, recall, and ROC curves (see Figure 1 for the ROC curve of different versions of the local alignment algorithm). To our knowledge, there is no baseline system with which to compare the performance of our approach.
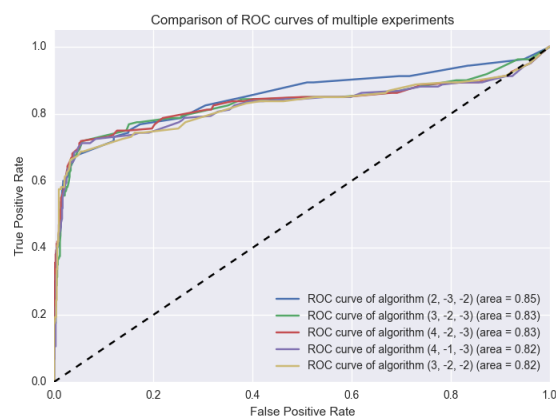


## Use Cases

On July 20, 2015, Wisconsin governor and 2016 presidential candidate Scott Walker signed into law a bill banning non-emergency abortions past the 19th week of pregnancy. Unsurprisingly, Walker's move garnered lots of media attention, but few journalists had enough time to figure out how many states have introduced similar legislation and where it originated. Using our prototype, we found 73 very similar bills that had been introduced around the country.

```
(1) legislative findings. the legislature finds that the best curr
ent evidence confirms: (a) pain receptors (unborn

----------------------------------MATCH----------------------------------

child's entire body  nociceptors) are present no later than  16
weeks after fertilization and nerves link these receptors to the
brain's thalamus and subcortical plate by no later than  20 week
s. (b) by  8 weeks after fertilization,  - - - - - - - - - - th
e unborn child reacts to stimuli that would be recognized as painf
ul if applied to an adult human, for example, by recoiling.  (c)
in the unborn child, application of  -  painful stimuli is associa
ted with significant increases in stress hormones known as the str
ess response. (d) subjection to  -  painful stimuli is associate
d with long-term harmful  neuro-developmental  effects, such as al
tered pain sensitivity and, possibly, emotional, behavioral, and l
earning disabilities later in life.  (e) for the purposes of surg
ery on unborn children, fetal anesthesia is routinely administered
and is associated with a decrease in stress hormones compared to
their level when painful stimuli is applied without  the anesthes
ia.  (f) the position, asserted by some medical experts, that the
unborn child is incapable of experiencing pain until a point late
r in pregnancy than  20 weeks after fertilization predominately r
ests on the assumption that the ability to experience pain depends
on the cerebral cortex and requires nerve connections between the
thalamus and the cortex. however, recent medical research and ana
lysis, especially since 2007, provides strong evidence for the con
clusion that a functioning cortex is not necessary to experience p
ain.  (g) substantial evidence indicates that children born missi
ng the bulk of the cerebral cortex, those with  hydraencephaly, n
evertheless experience pain.  (h) in adults, stimulation or ablat
ion of the cerebral cortex does not alter pain  perception while
stimulation or ablation of the thalamus does.  (i) substantial ev
idence indicates that structures used for pain processing in early
development differ from those of adults, using different neural e
lements available at specific times during development, such as th
e subcortical plate, to fulfill the role of pain processing.  - (
j) consequently, there is substantial medical  evidence that  an
unborn child

----------------------------------MATCH----------------------------------
```

```
journal of medicine, 317:1321-29 (1987). (8) pain receptors (nocic
eptors) are present throughout the unborn

----------------------------------MATCH----------------------------------

child's entire body  - - by no later than  sixteen weeks after
fertilization and nerves link these receptors to the brain's thala
mus and subcortical plate by no later than  twenty weeks.  (9) b
y  eight weeks after fertilization,  the unborn child reacts to t
ouch. after twenty weeks post-fertilization, the unborn child rea
cts to stimuli that would be recognized as painful if applied to a
n adult human, for example, by recoiling.  (10) in the unborn chi
ld, application of  such painful stimuli is associated with signi
ficant increases in stress hormones known as the stress response.
(11) subjection to  such painful stimuli is associated with lo
ng-term harmful  neurodeveolopmental  effects, such as altered pai
n sensitivity and, possibly, emotional, behavioral, and learning d
isabilities later in life.  (12) for the purposes of surgery on u
nborn children, fetal anesthesia is routinely administered and is
associated with a decrease in stress hormones compared to their le
vel when painful stimuli is applied without  such anesthesia.  (1
3) the position, asserted by some medical experts, that the unbor
n child is incapable of experiencing pain until a point later in p
regnancy than  twenty weeks after fertilization predominately res
ts on the assumption that the ability to experience pain depends o
n the cerebral cortex and requires nerve connections between the t
halamus and the cortex. however, recent medical research and analy
sis, especially since 2007, provides strong evidence for the concl
usion that a functioning cortex is not necessary to experience pai
n.  (14) substantial evidence indicates that children born missin
g the bulk of the cerebral cortex, those with  hydranencephaly, n
evertheless experience pain.  (15) in adults, stimulation or abla
tion of the cerebral cortex does not alter pain  perception, whil
e stimulation or ablation of the thalamus does.  (16) substantial
evidence indicates that structures used for pain processing in ea
rly development differ from those of adults, using different neura
l elements available at specific times during development, such as
the subcortical plate, to fulfill the role of pain processing.
(17) the position, asserted by some medical  experts, that  the
unborn child

----------------------------------MATCH----------------------------------
```
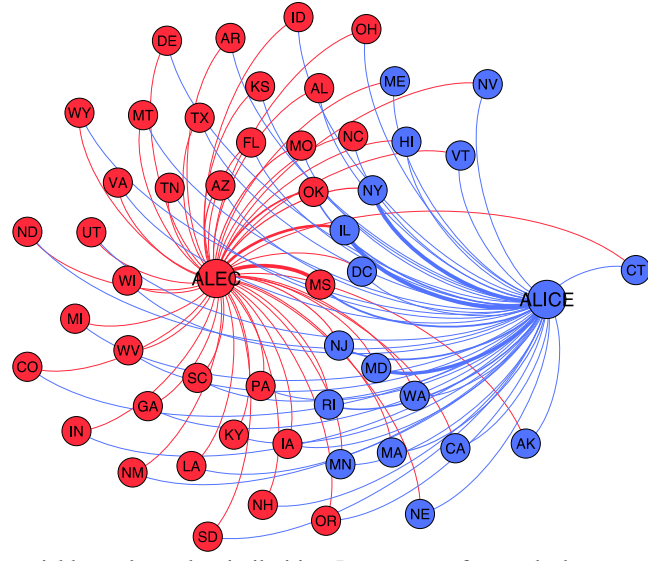
returns a list of documents that potentially match. The tool highlights similar sections in those documents, allowing the user

The screenshot below shows the tool's highest-rated match for Walker's bill. The left-hand side displays text from Wisconsin Senate Bill 179 (2015), and the right-hand side displays text from Louisiana Senate Bill 593 (2012). The highlighting shows that these sections match almost perfectly. Where differences exist, they are usually numbers versus spelling (e.g. "16" versus "sixteen") or section identifiers (e.g. "(b)" versus "(9)"). Thanks to the tool, we learned that similar bills had passed in Arkansas, Georgia, Idaho, Kansas, Oklahoma, Texas, and West Virginia; that similar bills were under consideration in Illinois, Iowa, Kentucky, Maryland, Oregon, South Carolina, and Virginia; and that similar bills had died in Florida, Michigan, Minnesota, Mississippi, and New Mexico.

In addition our tool allows researchers to do large-scale analysis of interest groups influence across state legislatures. We have collected more than 2500 bills drafted by interest groups and used the tool to detect the influence these groups have across states. For instance, we have analyzed the influence of the two main interest group organizations: ALEC, a conservative interest group, and ALICE, on the liberal side. The network below shows which states are more influenced by ALEC (in red) and which ones are more influenced by ALICE (in blue). The thicker the line, the more influence the group has on that state.

We believe both researchers and scholars will benefit highly from our tool. For researchers we will provide a downloadable data set that will allow them to analyze in greater depth how interest groups are influencing state politics. For journalists, we will provide an interactive app or search engine. In its current state, our tool allows the user to enter text of a bill and our system



to quickly evaluate the similarities. In a matter of seconds the user can look through a subset of bills to analyze whether a given bill is borrowing language from another state or an interest group. As a consequence, our tool will (i) increase transparency in state politics by revealing where bills come from and (ii) democratize the process of keeping state lawmakers accountable by enabling any individual to do this work.