The effects of transitions of power on the contents of municipal government websites

Markus Neumann Bruce Desmarais Hanna Wallach

May 18, 2017

Abstract

We explore the effect of transitions of power in municipal governments on the content of their websites. We hypothesize that when party control changes, city administrators modify the contents of their websites in order to fit the agenda of the new incumbent. To test this theory, we study cities in Indiana and Louisiana, two states in which all municipal elections are partisan and the parties of the candidates appear on the ballots. Snapshots of websites before and after transitions of power are acquired through the Wayback Machine. We apply statistical topic models based in latent dirichlet allocation, focusing on changes to the websites. We present results on both which topics see the greatest degree of change associated with transitions in city administrations, and how the topics modified differ with regard to political parties.

1 Introduction

2 Background

Grimmelikhuijsen (2010) run an experiment in which citizens are exposed to randomly selected levels of information about local government council minutes. They find a negative relationship between the information level and perceptions of competence in the local government. This raises an interesting question regarding whether citizens or more likely to participate when they perceive competence or when they perceive incompetence.

Wang, Bretschneider, and Gant (2005) present a widely cited 'model' for evaluating the accessibility of information on government websites. This is an important paper with which we should be familiar at a very detailed level as we use archived web content to assess the volume/accessibility of information provided by local governments.

Osman, Anouze, Irani, Al-Ayoubi, Lee, Balcı, Medeni, and Weerakkody (2014) is less relevant, but they develop a multi-item measure to predict the level of citizen satisfaction with e-government services.

Grimmelikhuijsen and Welch (2012) conduct an enormously relevant study. Insofar as we analyze what predicts openness of government websites, we will be replicating and building upon this study. They focus on Dutch municipal websites, and their approach is fairly limited in scope and highly manual (which we can compliment). For example, one of the dependent variables "Decision-making transparency," is measured "using a discrete (1/0) indicator for whether the underlying principles or reasons for local air pollution policies were given on the Web site."

Names	Year	Journal	Findings	Important?
Benedictis-	2016	JOP*	Regression discontinuity design. Democratic mayors spend more (but it is	Yes
ï,			unclear on what, not the typical Democratic issue-areas), issue more debt, pay	
Justin De Warshaw			more interest	
Christopher				
Caughey, Devin	2015	Working Paper	Regression discontinuity design. Partisan composition of state governments affects state policy liberalism (composite index for the areas of social welfare.	Somewhat
Warshaw,			taxation, labor, civil rights, womens rights, moral legislation, family planning,	
Christopher Xu, Yiqing			environment).	
Einstein,	2015	Urban Affairs Re-	Cities with more Democratic citizens spend more; more progressive (rather	Somewhat
Katherine		view	than regressive) forms of taxation; pursue intergov. aid more; spend more on	
Levine			police, fire, parks & recreation	
Kogan, Vladimir				
Einstein,	2015	Working Paper	Survey of 72 mayors. Unlike Republican mayors, roughly half of Democrats	Somewhat
Katherine			seem to agree that cities should aim to reduce inequality. Democratic mayors	
Levine			also seem to favor redistribution to accomplish that goal.	
Glick, David				
M.				
Kiewiet, D Roderick	2014	Annual Review	City budgets have been severeley constrained since the Great Recession. Spending has thus decreased in general. Lack of funds means that there is	Somewhat
Mccubbins,			not much discretion for partisanship.	
Mathew D				
\circ	2014	APSR*	Cities are responsive (taxes, expenditures, regressiveness of taxation) to citi-	Yes
Chris War-			zens' conservatism/liberalism. Partisan elections do not make cities more or	
shaw, Christopher			less responsive.	
Guillamón,	2013	European Jour-	Police spending in Spain. Conservative parties spend more on police. Spend-	Yes
Ma Dolores		nal of Law and	ing is higher before elections. Also contains a useful overview of the literature.	
Bastida,		Economics		
Francisco				
Benito,				
Bernardino				

Gerber, Elisabeth R.	2013	Cityscape	Partisanship of both citizens and elected city officials separately affect climate policy.	Yes
Names	Year	Journal	Findings	Important?
Solé-Ollé, Albert	2013	Journal of Urban Economics	Spanish cities. The authors "employ a regression discontinuity design to document that cities controlled by left-wing parties convert much less land from	Yes
Viladecans-			rural to urban uses than is the case in similar cities con- trolled by the right".	
Marsal, Elisabet			Partisanship might also affect housing construction and price growth.	
Gerber,	2011	AJPS	Regression discontinuity design. Democratic mayors spend less on public	Yes
Elisabeth R Honkins			\supset	
Daniel J.				
Trounstine,	2010	Annual Review	Race and ethnicity in local elections (not relevant to us). Partisan elections	Somewhat
Jessica			have higher turnout; non-partisan elections still tend to have some partisanship in them because voters learn about party of candidates from media. Non-	
			partisan elections favor Republicans/upper class. Mixed evidence for whether	
,		,		,
Palus, Chris-	2010	State and Lo-	_	Somewhat
tine Kelleher		cal Government	ing in five areas: (1) community development, housing, and conservation, (2)	
		Review	health and human services, (3) culture, the arts, and recreation, (4) environmental programs, and (5) transportation	
Ferreira,	2009	The Quarterly	esults for spending and city gov. size	Yes
Fernando		Journal of Eco-	with regard to mayor partisanship.	
Gyourko,		nomics		
Joseph				
Ansolabehere,	2006	Scandinavian Jour-	Despite the journal, this is about the U.S. The important finding (for us) is	Somewhat
Stephen		nal of Economics	the fact that counties whose government is controlled by the same party as	
Snyder,			the state government, receive more funding (county's share of state transfers,	
James M.			normalized by county pop.) from the state.	
Murphy,	2002	Annual Review	cites papers written a hundred years	No
Russell D.			ago. Also exclusively about larger cities.	

There doesn't seem to be much literature on transitions of party control, if anything, that question is mostly phrased with regard to political representation. However, if we want to tie our paper to a larger theory, we could go with dynamic representation. Under dynamic representation, policy-makers are responsive to trends in citizens opinions, which mainly manifest/become apparent through election results, especially when incumbents are voted out of office. This fits our topic quite well. Also, of the few papers that do exist on the effects of partisan transitions, virtually all use regression discontinuity designs.

3 Data

The General Services Administration (GSA) maintains all .gov addresses, and provides a complete¹ list of all such domains to the public through GitHub². This list is updated once per month—we rely on the version released on January 16, 2017. The data from the GSA contains the following variables: One, domain name, specifically, the all-uppercase version of domain and top-level domain (for example, 'ABERDEENMD.GOV'). Two, the type of government entity to which the domain is registered, such as city, county, federal agency, etc. Three, for federal agencies, the name is specified. Finally, the city in which the domain is registered, is noted.

Here, we focus only on cities. As a first step, we use a webdriver-controlled browser (Firefox/Selenium/Geckodriver) to test whether all of the city websites actually work. Of the 2425 domains listed by the GSA as cities, 292 are not accessible. Furthermore, the .gov domain, as registered at the GSA, is frequently not the website a city actually uses. In many cases, these sites redirect to another address, sometimes not a .gov domain (in this case, we simply use this domain). We record these URLs, as they are required to retrieve the images websites stored in the Wayback Machine (WbM).

In order to provide an overview of our coverage (as not all cities, towns and villages use .gov addresses), we merge this list with U.S. Census data³. Here, several limitations in the GSA data need to be accounted for: One, even though the GSA nominally separates websites of cities and counties, some of the domains categorized as cities actually belong to counties. The same is true for townships and boroughs. Ergo, we eliminate all websites belonging to these three types of entities by hand. Furthermore, the city name, as given by the GSA, refers to the city in which the domain is registered, which is not necessarily equivalent to the city the website serves. In many cases, a website of a larger city may be registered to one of its subdivisions (for example, the website of New York is registered to Brooklyn), or vice versa (for example, the website of Homecroftin, a small town within Indianapolis, is registered to the city as a whole). Consequently we fix mismatches between websites and cities manually. Finally, a number of cities are simply misspelled, which we also correct by hand.

¹Domains used for testing and internal programs are excluded.

²https://github.com/GSA/data/tree/gh-pages/dotgov-domains

³http://www2.census.gov/programs-surveys/popest/datasets/2010-2015/cities/totals/sub-est2015_all.csv

After the counties, townships and cities that cannot be matched to the Census data⁴ and duplicate websites (some cities have more than one website) are removed, 1813 domains/cities remain.

These cities contain 90,616,865 people, and thus about 28% of the U.S. population (see figure 1).

We use the resulting list of websites to access their copies stored in the Internet Archive's Way-back Machine. To this end, we rely on the Ruby Gem 'Wayback Machine Downloader' (WbMD). We supply the URL that each .gov website redirects to to the WbMD, which then downloads every file present in the WbM from a snapshot in October 2016, or, if not available, as soon as possible after this point.

<Note: We have not actually done this last step for all websites (however, the R script which runs the Ruby package is already set up to do so once we need to). Instead 10 websites were randomly sampled from an older version of the GSA list, which still contained counties and townships, which is why one of the 10 websites is from Dutchess County, NY.>

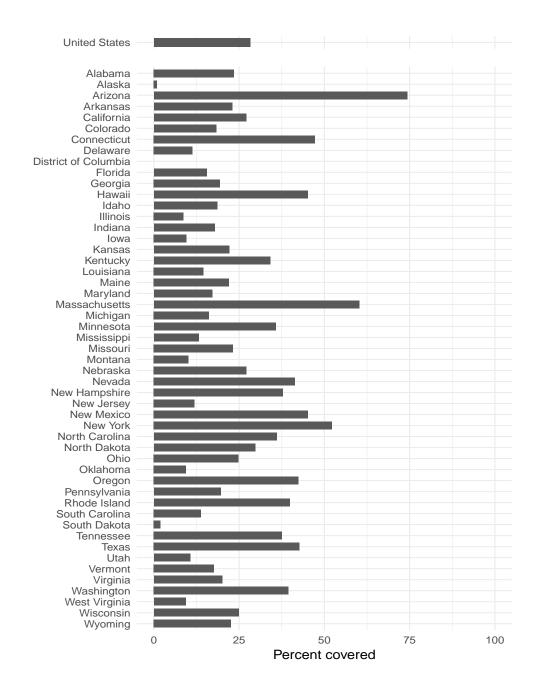
3.1 Indiana City Websites

It would be fine to focus on Indiana as a case. First, we need to answer some preliminary questions about the data.

- 1. For what percentage and number of IN cities can we find data from the WBM?
- 2. For how many election cycles can we find political leadership data for these matched cities?
- 3. In what number and percentage of cities is the local leadership majority Republican?
- 4. Relatedly, in a typical election cycle, for how many cities do we see a transition in party leadership (i.e., a shift from majority D (R) to majority (R) D).
- 1. 30 cities, with a combined population of 1,180,435. However, since only cities (as opposed to towns and villages) hold mayoral elections, only 16 of these, with a combined population of 1,094,383 can be matched to the election data.
- 2. 2015, 2011, 2007, 2003.
- 3. Of the 16 cities, 7 have Republican mayors after the 2015 elections.
- 4. In 6 cases, a shift of party control occurs, with 4 of these being Republican -> Democratic.

⁴There are five cities that are not contained in the Census data

⁵https://github.com/hartator/wayback-machine-downloader



3.2 Research Design

1. Corpus:

E11.4	F
Filetype	Frequency
	47652
pdf	9486
jpg	4342
html	3492
aspx	2832
png	2358
gif	1296
JPG	545
css	385
1	358
js	344
htm	264
docx	202
dwf	181
doc	168
asp	161
PDF	103
php	83
egov	54
cfm	45
4	42
2	35
com	31
0	29
3	26

Table 1: The 25 most common file types in scraped websites

- (a) Last snapshots before the election (November 3, 2015 in Indiana; tbd. in Louisiana (probably February))
- (b) First snapshot that is at least 2 months after the new government's inauguration (which is in January for Indiana, May for Louisiana)

2. Preprocessing:

- (a) restrict corpus to:
 - i. documents belonging to cities in which a change of power occurred
 - ii. documents that were added, deleted or changed between the two snapshots
- (b) words to lowercase

- (c) remove punctuation
- (d) stemming (Porter stemming algorithm?)
- (e) Remove stop words (regular list of stop words is enough, since we use an asymmetric prior)
- 3. Apply Grimmer's expressed agenda model to the corpus
 - (a) Asymmetric prior
 - (b) Each document can have only one topic (in contrast to the author-topic model)
 - (c) Cities i = 1, ..., n = 15
 - (d) Topic k(k = 1, ..., K)
 - (e) Documents $j(j = 1, ..., D_i)$ from city i
 - (f) Party covariate in the prior, where the deleted and unmodified documents are coded as from the first, and the added and modified documents from the second party

4. Results

- (a) Label topics using Grimmer's automatic cluster labeling method, based on most commonly used words in documents belonging to topic
- (b) Evaluate topics

Validation:

- Do the above for cities in which no change of power occurred.
- Check whether there is higher than average turnover around the new year by comparing changes to non-election years (and also Louisiana, where elections are later).
- Check how long documents stay on websites on average. Use websites with a lot of snapshots for this (these exist for both small and large cities).

Problem with using this model: Grimmer's expressed agenda model uses Senators as the actors. Senators is also who he is substantively interested in. For us, the equivalent to Senators is cities. However, we care about parties, not cities.

3.3 Survival model

The existence of individual documents on municipal government websites can be though of as a survival process. No document stays on a website forever, and it appears to be a reasonable assumption that as documents get older and thus less relevant, they get replaced. The factors determining the steepness of the survival curve are the topic - fire safety regulations likely stay up longer than a bulletin on the annual spring banquet - and the change of party control after an election.

- H1: The older a document, the more likely it is to be removed.
- S(t) has a downward slope. Admittedly, this is almost impossible not to be true. Also, test proportional, rising and falling hazard models.
 - H2: Documents pertaining to administrative matters are less likely to be removed.

Introduce a categorical variable for the top 10(?) topics. A negative coefficient for administrative topics would support this hypothesis.

H3: Documents introduced by the opposing party are more likely to be removed.

Introduce two variables into the survival model: One variable indicating which party has introduced a document, and a time-varying variable describing which party is currently in government. The hypothesis is tested through an interaction term between the two.

H4a: Democrats are more likely to remove documents with topics pertaining to private enterprise, private schools.

Interaction term between party in power and categorical topic variable.

H4b: Republicans are more likely to remove documents with topics pertaining to social justice, equality, taxes, public schools, etc.

Interaction term between party in power and categorical topic variable.

H5: In line with their commitment to small government, Republicans are more likely than Democrats to remove documents.

Party in power variable.

This model will take up a lot of degrees of freedom. The rarity of snapshots for some cities might be a problem. Documents being changed and being removed can be modeled as competing risks.

Y =Party that introduced the document

+ Party that is currently in power

$$+$$
 Topic 1, topic 2, ..., topic k (1)

- + Party that is currently in power \times Topic 1, topic 2, ..., topic k
- + Days since start of mayoral term (control)

3.4 Topic modeling

Note: this chapter is mostly a wordy and less coherent version of the above.

We hypothesize that a change in leadership from one party to the other will lead to a change in website content because the two parties have different agendas. Democrats have a predilection towards policies that promote social and economic equality, whereas Republicans like to emphasize small government as well as law and order. Documents uploaded to city websites are expected to be a reflection of these preferences.

The Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003) is the most commonly used topic model. However, it is unable to account for the existence of two parties with very different policy agendas, translating to different preferred topics. There are two types of extensions to the LDA that fit our subject much better - the structural topic model, and the author-topic model.

The structural topic model, developed by Roberts, Stewart, and Airoldi (2016) allows researchers to model a corpus as a function of metadata associated with its documents. Specifically, topic prevalence (the proportion of a document made up by a topic) and topical content (the rate at which words figure into a topic) are contingent on a set of covariates. In our case, the two most important covariates are (1) city and (2) authoring party (operationalized by whether a document was present before a change of power, or introduced afterwards). Furthermore, the population size of the city should be a predictor for both the number and kind of problems it faces, which thus need to be addressed on its website. Furthermore, city size also serves as an estimate for the budget and technical capacities of its staff in charge of maintaining the website⁶. Further demographic as well as economic data might also be useful to differentiate cities from one another. If we model the differences between cities properly, we might not have to/should not include city as a (categorical) variable, because it would probably interfere with these more meaningful covariates.

The author-topic model (Rosen-Zvi, Griffiths, Steyvers, and Smyth 2004) would allow us to capture the fact that different authors have different topical preferences. Unfortunately, we have two types of 'authors' - cities, and parties. Given the largely divergent administrative needs of different types of cities, we would likely have to treat cities as the author. This would require us to capture the partisan authorship of different documents entirely on the basis of sub-sets of the website data - changed, added, or deleted documents. (Note: In the papers on author-topic models, the intention is often to analyze scientific articles. These articles are often co-authored. Would it be possible to have BOTH cities and parties as authors, so that a specific version of a website would then be 'co-authored' by its city and party?)

The critical element in this analysis is to accurately attribute authorship of documents to either party. Despite possible changes to websites due to a leadership transition, large parts of the content carry over. This means that unless the successor government decides to delete everything, some of the existing documents will be preserved, and in the model, also attributed to the new 'author'. But the reverse is not possible, because the predecessor government can't choose to retain documents

⁶Although this relationship is not exactly deterministic - when looking through .gov websites manually, I've noticed that a lot of websites of (presumably wealthy) towns of only a few thousand citizens often have extremely well-kept websites

from the future. This is a very important point for municipal websites. We should investigate the possibility of modeling only the changes—documents that change, documents that are deleted, and documents that are added.

Labeling newly added documents after a change of power is quite simple. As far as older documents are concerned, we would have to operate under the assumption that the incumbent didn't keep his or her successor's document's on the website for four years. One problem here is the fact that the incumbent would have all the administrative topics assigned to them, simply because they have to have those on their website.

If we really do end up getting swamped with administrative terms in our topic models (and it does kind of look like that at the moment), we might be able to separate the signal from the noise by running a preparatory LDA once and using its results to create a new, corpus-specific list of stop words. After that, we run the actual model. This way, politically charged terms and topics, which likely are not as common, but present nevertheless, should be able to rise to the surface. It might be possible to refine this process by running an exploratory model on website data from cities in which party control never changes, and the incumbent always wins by large margins. 'Safe' cities like this should have fairly homogeneous populations, with little need for the incumbent to play politics on the municipal website. Hence, these websites should be filled with purely administrative content.

The use of asymmetric priors (Wallach, Mimno, and Mccallum 2009) over the document topic distribution - i.e. the assumption that some topics, such as administrative content, are inherently more common - may be a more elegant way of dealing with this issue.

Another intervening factor is that for cities in Indiana, mayoral terms begin in January. Since a lot of clerical and administrative tasks tend to be year-specific, work tends to pile up around the new year. Thus it is possible that a spike in newly added documents is not due to a change in party control, but owed to a seasonal increase in activity. We can test for this by comparing election years to non-election years. Furthermore, since in Louisiana, mayors take office in May, we have another point of comparison.

Furthermore, if we only investigate cities in which control of government changes from one party to another, we may overestimate its effect. Not only does a transition in party control occur, but the person in charge also changes. Parties are fairly homogeneous, so that two mayors from the same party may have very different policy preferences and managerial styles. To remedy this problem, we [could] utilize matching, pairing our cases with similar cities in which the incumbent does not run for re-election, but party control stays the same nevertheless.

⁷Probably a safe assumption. However, we could, and probably should test how long documents tend to stay on a city website. Simple descriptive statistics (for example density plots) on the length of existance should likely be sufficient. If we want to be really fancy about it, we could create a duration model, with document topics as features. This would allow us to measure whether some documents tend to remain longer based on their topic (i.e. fire regulations are probably going to stay up longer than notes on a specific council meeting).

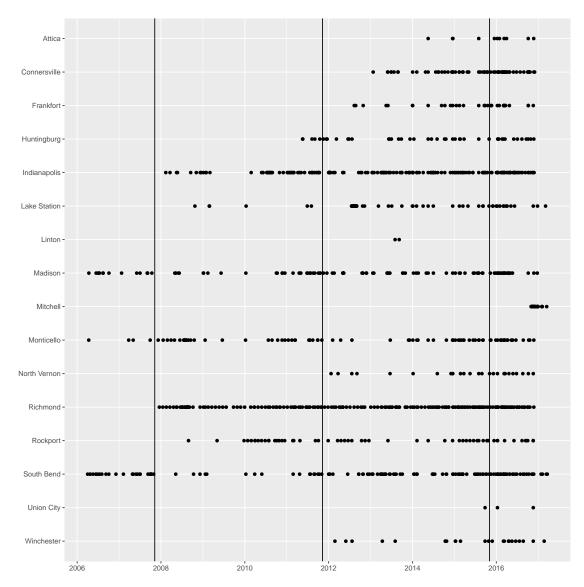
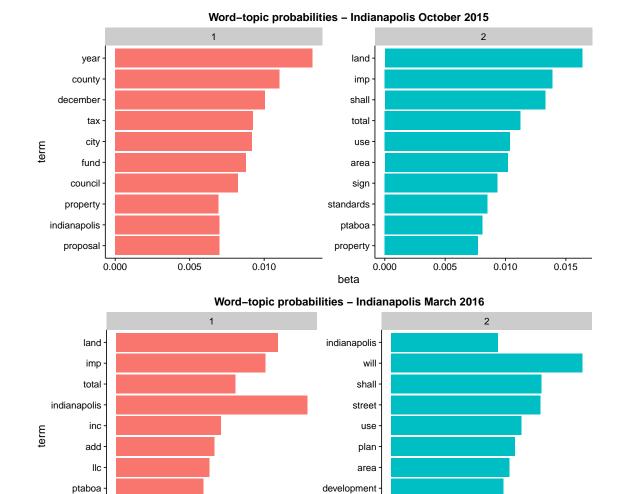


Figure 1: Dates of Wayback Machine snapshots. The vertical lines are municipal elections.

References

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003. ISSN 15324435. doi: 10.1162/jmlr.2003.3.4-5.993.



Stephan G Grimmelikhuijsen. Transparency of public decision-making: Towards trust in local government? *Policy & Internet*, 2(1):5–35, 2010.

0.03

school

building

beta

0.000

0.002

0.004

0.006

0.008

construction

ave

0.00

0.01

0.02

Stephan G Grimmelikhuijsen and Eric W Welch. Developing and testing a theoretical framework for computer-mediated transparency of local governments. *Public administration review*, 72(4): 562–571, 2012.

Ibrahim H Osman, Abdel Latef Anouze, Zahir Irani, Baydaa Al-Ayoubi, Habin Lee, Asım Balcı,

- Tunç D Medeni, and Vishanth Weerakkody. Cobra framework to evaluate e-government services: A citizen-centric perspective. *Government Information Quarterly*, 31(2):243–256, 2014.
- Margaret E Roberts, Brandon M Stewart, and Edoardo M. Airoldi. A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, pages 1–49, 2016. ISSN 0162-1459. doi: 10.1080/01621459.2016.1141684. URL http://www.tandfonline.com/doi/full/10.1080/01621459.2016.1141684.
- M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth. The author-topic model for authors and documents. *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 487–494, 2004. ISSN 01689002. doi: 10.1016/j.nima.2010.11.062. URL http://portal.acm.org/citation.cfm?id=1036902.
- Hanna M Wallach, David Mimno, and Andrew Mccallum. Rethinking LDA: Why Priors Matter. *Advances in neural information processing systems*, 2009.
- Lili Wang, Stuart Bretschneider, and Jon Gant. Evaluating web-based e-government services with a citizen-centric approach. In *System Sciences*, 2005. *HICSS'05*. *Proceedings of the 38th Annual Hawaii International Conference on*, pages 129b–129b. Ieee, 2005.

Website	Files	Size (kB)
attica-in.gov	1417	61988
bedford.in.us	560	57628
cityofboonvilleindiana.com	110	9848
connersvillecommunity.com	569	170684
frankfort-in.gov	2652	205368
mitchell-in.com	50	3036
warsaw.in.gov	2117	298264
www.batesvilleindiana.us	1	96
www.bernein.com	755	60836
www.bloomington.in.gov	1	52
www.brazil.in.gov	1	36
www.carmel.in.gov	8727	2267604
www.ci.auburn.in.us	1025	178496
www.ci.bluffton.in.us	1	132
www.cityoffortwayne.org	4378	2134540
www.cityofhobart.org	2463	718312
www.cityofrisingsun.com	3311	991736
www.cityofrockport-in.gov	98	11836
www.clintonco.com	524	134344
www.elkhartindiana.org	2345	1132612
www.elwoodcity-in.org	1	48
www.evansvillegov.org	11844	6339388
www.gary.in.us	1227	371480
www.huntingburg-in.gov	2496	388188
www.indy.gov	9675	5717888
www.jasperindiana.gov	4013	561212
www.lakestation-in.gov	2	40
www.linton-in.gov	1	32
www.madison-in.gov	1848	528200
www.martinsville.in.gov	1463	46544
www.monticelloin.gov	753	33352
www.newhavenin.org	626	84176
www.northvernon-in.gov	403	271452
www.richmondindiana.gov	1042	250028
www.southbendin.gov	4749	1252624
www.tiptongov.com	1151	130404
www.unioncity-in.gov	2	16
www.winchester-in.gov	2480	364368

Table 2: Number of files and size of websites

	tf
tax	176324
date	98949
due	97192
amt	96382
town	86726
value	81119
total	70825
parcel	63589
county	56201
market	51758
east	51357
full	51199
nrth	50566
book	50306
deed	50231
bill	49719
acres	48935
acct	44871
csd	43792
owners	43510
res	41803
family	36883
fire	35156
school	33685
name	30382
red	30362
taxable	30248
hook	29902
homestead	29287
outside	28593

Table 3: Top term frequencies for 10 test websites

City	DemVotes	RepVotes	Winner	Change	Pop15	url
Attica		187	Republican	0	3117	https://attica-in.gov/
Connersville	1005	995	Democratic	1	13010	http://connersvillecommunity.com/
Frankfort		1748	Republican	0	16060	http://frankfort-in.gov/
Huntingburg	447	793	Republican	0	6035	http://www.huntingburg-in.gov/
Indianapolis	92830	56661	Democratic	1	862781	http://www.indy.gov
Lake Station	1483	227	Democratic	0	12054	http://www.lakestation-in.gov/
Linton	785	692	Democratic	0	5284	http://www.linton-in.gov/
Madison	1192	1915	Republican	0	12040	http://www.madison-in.gov/
Mitchell	229	495	Republican	1	4252	http://mitchell-in.com/
Monticello	0		Democratic	0	5322	http://www.monticelloin.gov/
North Vernon	679	697	Republican	1	6619	http://www.northvernon-in.gov/
Richmond	3421	2731	Democratic	0	35854	http://www.richmondindiana.gov/
Rockport	286	272	Democratic	1	2223	http://www.cityofrockport-in.gov/
South Bend	8515	2074	Democratic	0	101516	https://www.southbendin.gov/
Union City	338	440	Republican	0	3447	http://www.unioncity-in.gov/
Winchester	606	524	Democratic	1	4769	http://www.winchester-in.gov/

Table 4

File type	oct15	jan16	feb16	mar16
.pdf	179	183	111	197
.aspx	99	112	116	76
.jpg	22	24	17	12
.gif	3	5	1	5
.JPG	3	2	1	5
.png	3	7	2	1
.asmx	2	1	1	5
.PDF	1	1	1	2

Table 5: File types of Indy.gov at different points in time