Simon School Collab. Techical Report
Deriving Syntactical Errors in Noisy Form 10-K's

Chester Holtz
Department of Computer Science
University of Rochester
Rochester, NY 14627, USA

1. INTRODUCTION

The aim of this project is to approach detection of grammatical and spelling errors in form 10-K's from the perspective of a computer scientist. For the purposes of this report, we will give a brief overview of the methods taken without delving into great depth. A more rigorous explanation of the algorithms implemented can be found in cited research, or a future work.

We define *spelling errors* as not only words that do not exist in the English language (non-word errors), but also written words whose spelling differs from that of the word intended by the author (real-word errors) - i.e., if the author writes *"I went two the park"*, the word *two* should be classified as a misspelled word since the author clearly meant to use the word *"to"*. *Grammatical errors* are errors which cause a sentence to not comply with English grammar rules. Developing a system to detect grammatical errors in a text requires that the program understands the contextual information conveyed by sentences in the text.

Form 10-K's are annual reports that provide the SEC a comprehensive summary of a company's financial standings. These forms contain information such as the company's history, organizational structure, executive salary, equity, subsidiaries, and audited financial statements. Form 10-K's are often filed as XBRL documents before being uploaded to the EDGAR central database.

The XBRL file format is the current standard for expressing semantic meaning for business reports. An XBRL document is written in an XML-based language and consists of an XBRL *instance*, containing a set of business facts, and a set of *taxonomies* which detail metadata and relationships between facts. Like XML, XBRL documents contain structured trees, with the root nodes of the trees expressed as $< xbrl >$ tags. In a single document, there may contain multiple trees.

XBRL Instances
Within an XBRL instance, there may contain the following:

(1) Business facts are each assigned to a context and contain
  (a) Items containing a single value
  (b) Tuples holding multiple values, or nested tuples
(2) Contexts define the entity (a company or individual) a fact refers to, the period of time the fact is relevant, and an optional scenario. Scenarios provide further contextual information about the facts, such as whether the business values reported are actual, projected, or budgeted.
(3) Units define the units used by numeric or fractional facts within the document, such as USD.
(4) Footnotes associate one or more facts with some content.
(5) References to XBRL taxonomies, typically through schema references. It is also possible to link directly to a linkbase.

XBRL Taxonomy
A Taxonomy is a collection of schemas and linkbases to external XBRL compliant documents.

(1) Schemas define item and tuple concepts
(2) Linkbases are collections of links

When filed with SEC EDGAR, each form is subject to a publishing process where the document is converted to EDGAR-compliant HTML. We observe that this process is far from perfect. Since many of the forms are uploaded in XBRL format and undergo the conversion process, the resulting output is very noisy and difficult to parse. Our goal is to implement a system to extract financial facts from these filings as well as any natural language description and to apply relevant NLP techniques to derive syntactical errors in the human-written component.

2. RELATED WORK

In this section, we summarize related work in the fields of grammatical and spelling error detection. For a more detailed review, the readers can refer to Leacock et al.[6]. In addition to reviewing relavent research, we will go over the popular tools used for HTML parsing and syntactic error detection.

Popular libraries exist for parsing HTML and XML. For example, html.parser and Beautiful Soup are Python libraries that can derive a parse tree from scraped HTML which can then be used to extract data from the webpage. We found Beautiful Soup to be a viable approach with some caveats. Our primary concern was that we found Beautiful Soup's clean function to be too slow - even with the lxml parser backend. Additionally, we found the implementation to be less than robust on the form 10-K's and that the system became confused with the malformed generated HTML. In this instance, we implemented our own HTML to markdown converter based off of a modified version of Aaron Swartz's html2text project. To improve the preformance of the script on the malformed HTML, we ran the forms through Beautiful Soup's `prettify` system before parsing.

To derive syntactical errors from the parsed forms, we identified prior research to derive an approach. Our primary goal is to identify spelling mistakes, so we currently ignore spelling correction. Traditional approaches to the spelling classification problem center around creating an exhaustive lookup table of unique English

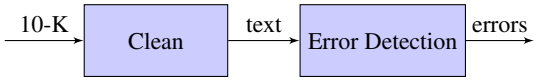
words. If a word does not exist in this table, it is classified as a misspelled word.

Examples where this model does not perform optimally are with words that are real words, but not spelled as the author intended them to be spelled. These errors include typographical errors (i.e. the word *three* being misspelled as *there*), cognitive errors (the word *too* being misspelled as *two*), and words which are real words, but do not exist in our wordlist.

Current approaches to solving this problem involve training an HMM with a trusted corpus of natural language. In particular, HMMs that model spelling errors by examining the most likely sequence of part-of-speech tags have been shown to be successful by Li et al. and Fossati, Di Eugenio[3]. Both Li et al. and Fossati, Di Eugenio trained their models on corpora derived from Wikipedia articles. Other approaches involve utilizing preexisting libraries. One example of an effective library is PyEnchant. The PyEnchant package is a library based on the libEnchant C library which applies Lawrence Philips's Metaphone algorithm to first organize words into similar sounding groups before applying some hash function to index the words. When a word does not exist in the dictionary, libEnchant will derive the most similar sounding word.

Detecting grammatical errors is more difficult and is still an active research area. Syntax-based checking (Jensen et al.[5]) involves constructing a complete parse tree of a sentence-based on a formal grammar. Since there exists no complete grammar for the English language, this model is unable to fully classify general sentences. Current approaches revolve around application of machine learning and statistics to detect grammatical errors. Han et al.[4] showed that the classification approach to error correction is advantageous by making use of powerful machine learning algorithms in conjunction with arbitrary features from the sentence context. Typical features include surrounding N-grams, POS tags, chunks, etc. The SMT approach is also popular. For example, the work in (Desilets and Hermet, 2009) uses simple round-trip translation with a standard SMT system to correct grammatical errors. Park and Levy[11] propose a noisy channel model for error correction and relies on a bigram language model to find grammatical corrections. Lee and Seneff[7] describe a lattice-based correction system. In our own approach, we rely on rule-based checking to determine if a sentence matches a set of rules. We then construct a lattice-based correction system and do a standard beam search to find the most descriptive error.

3. ERROR FRAMEWORK AND IMPLEMENTATION



HTML & XML Parsing

Parsing HTML and XML derivatives is known to be a difficult problem. New programmers who want to extract information from an HTML or XML document often attempt to do so using regular expressions. This is rarely a good idea. HTML and XML are irregular languages and regexes are inadequate for this purpose[12]. One explanation derives from the equivalence of regular expressions to finite automata coupled with the fact that the ability to nest tags in both XML and HTML to arbitrary depth requires some kind of memory to parse. In particular, the HTML generated by the EDGAR system is generated programmatically from already malformed XBRL code. Our system must be robust and ignore extraneous tags - extracting only the data that a business application needs to process.

We wrote an HTML parse library based on a modified version of Aaron Swartz's html2text script to construct a finite state machine to match open and closing tags. Each time we match an open tag, we enter its context and recursively continue until the context is exited with a closed tag, at which time we remove both tags. At this time, since we have not perfected an approach to dealing with the malformed HTML, we initially run each through through Beautiful Soup's `prettify` function, which does slow down the overall runtime, but not to the extent that `getText` does.

Spelling Error Detection

For this component of the framework, we developed two systems. One dictionary-based and another HMM-based. Currently, we employ the pure dictionary-based implementation. We just want to check if a word exists in our wordlist. If the word does, then we can say it is spelled correctly. For this project we compiled a custom word list. Since there exists some vocabulary in Form 10-K's not defined in established dictionaries, we repeatedly sample words from a training set of 10,000 forms and manually check them for use in Wikipedia articles. For the dictionary component of the first model, we apply the murmurhash3 hash function to each word in our dictionary.txt file and store the resulting key pairs in a dictionary. To reduce overhead cycles and support easy multiprocessing, we can precompute the dictionary and pickle it to a file. Given an arbitrary word, we can check to see if it is spelled correctly by hashing the word and checking the associated key in the dictionary. If the position is not filled, we classify this word as misspelled.

For our second method, we implemented a modified version of the approach taken by Fossati and Di Eugenio[3]. In our process - which is equivalent to that of Fossati and Di Eugenio's (see their paper for details) - we alter the length of the N-Gram to be 3 and re-implement their Metaphone algorithm to the second iteration[10].

Once implemented, we found that doing Viterbi on the HMM for each sentence is not feasible without extended hardware. Since Viterbi is difficult to parallelize, we cannot improve the runtime of $O(K^2n)$. Although we found for individual sentences, the second approach preformed better on phrases involving misspelled words that remained actual words, we chose to do the analysis using the classical dictionary approach.

Grammatical Error Detection

For the grammatical error detection, we employed the $language\_check$ api. The $language\_check$ api is a wrapper for the LanguageTool open source project and supports functionality for individual sentence classification. Given a sentence, the api returns a set of grammatical subrules that the sentence violates. In LT, canonical grammatical rules are represented by a set of subrules. We note that the approach taken by Naber is the current fastest solution we have tested, but still quite effective - achieving 91% precision, 84% accuracy, and 51% recall[9]. Verifying if the argument is grammatically correct is not as simple as checking if this set is empty. There are certain rules we want to ignore. For example, whitespace rule violations can be ignored. This rule includes all subrules involving whitespace - i.e., the use of

whitespace before comma and before/after parentheses (denoted WP_C, WP_P). We can represent these rules as a directed graph and traverse this graph with an efficient beam search to derive the most significant errors. If there exists a significant error in the top five results, we mark the sentence is grammatically incorrect. Some of the errors we keep track of are:

(1) agreement *"They **is** my favorite Canadian authors"*
(2) comma errors *"She jogged for 30 minutes**,** and walked for 20 minutes."*
(3) similar words *"She is better **then** me"*
(4) parallelism *"He likes to swim and **running**"*

In addition to the LT rules, we also explicitly check the following:

(1) sentence length exceeds 20 words
(2) a vs an
(3) whitespace post punctuation (.,?!:;)
(4) repeated words

4. SAMPLE RESULTS

Sample results for a test xbrl document are included in **Appendix A.**. The wordlist is included in the project's github page `https://github.com/Choltz95/warner_statement_project`.

5. CONCLUSION & FUTURE WORK

The results of these experiments are promising, and represent a good starting point for future research. Among the others, there are several points that would be worthy of further investigation and improvement:

(1) Continue to extend coverage of the wordlist
(2) Improve the robustness of the form parsing system
(3) Explore other methods of grammatical error detection

6. REFERENCES

[1] Bartley, Chen, Taylor. *A Comparison of XBRL Filings to Corporate 10-Ks - Evidence from the Voluntary Filing Program*.

[2] Bartley, Chen, Taylor. *A Comparison of XBRL Filings to Corporate 10-Ks - Evidence from the Voluntary Filing Program*.

[3] Fossat, Di Eugenio. *I saw TREE trees in the park: How to correct real-word spelling mistakes*. Chicago, IL,: University of Illinois, 2008.

[4] Han, Chodorow, Leacock. *Detecting Errors in English Article Usage by Non-native Speakers*, Journal of Natural Language Engineering., 2006.

[5] Jensen, Heidorn, Miller, Ravin. *Parse Fitting and Prose Fixing: Getting a Hold on Ill-formedness*, AM. J. Computational Linguistics, 1983.

[6] Leacock, Chodorow, Gamon, Tetreault. *Automated Grammatical Error Detection for Language Learners*. Synthesis Lectures on Human Language Technologies, 2010.

[7] Lee, Seneff. *Automatic Grammar Correction for Second-Language Learners*. Cambridge, MA, MIT , 2006.

[8] Li, Duan, Zhai. *CloudSpeller: Query Spelling Correction by Using a Unified Hidden Markov Model with Web-scale Resources*. Chicago, IL,: University of Illinois, 2008.

[9] Naber. *A Rule-Based Style and Grammar Checker*. Technische, Fakultt,: Universitt Bielefeld, 2003.

[10] Phillips. *Lawrence Philips' Metaphone Algorithm*. `http://aspell.net/metaphone/`

[11] Park, Levey. *Automated Whole Sentence Grammar Correction Using a Noisy Channel Mode*. San Diego, CA,: University of California, San Diego, 20011.

[12] Sipser. *Introduction to the Theory of Computation*, 3rd ed., 2012.

[13] Swartz. *html2text (The Asciinator)*. `https://github.com/aaronsw/html2text`

Appendix A.

raw form data

```
<div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="center"><font style="DISPLAY: inline; FONT-SIZE: 18pt; FONT-FAMILY: Times New Roman"><strong>PART
  I</strong></font></div>
<div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="center">
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="center"><br></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="center"><font style="DISPLAY: inline; FONT-SIZE: 12pt; FONT-FAMILY: Times New Roman"><strong>ITEM
    1.</strong></font><font id="TAB2" style="COLOR: black; LETTER-SPACING: 27pt"> <font id="TAB2"
style="LETTER-SPACING: 9pt">   </font> </font><font style="DISPLAY: inline; FONT-SIZE:
12pt; FONT-FAMILY: Times New Roman"><strong>BUSINESS</strong></font></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><br></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT:
0pt" align="justify"><font style="DISPLAY: inline; FONT-SIZE: 12pt; FONT-FAMILY: Times New
Roman"><strong><u>GENERAL</u></strong></font></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><br></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 36pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT:
0pt" align="justify"><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY: Times New
Roman"><strong><em>OVERVIEW
    AND DESCRIPTION OF SUBSIDIARIES</em></strong></font></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><br></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 18pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY: Times New Roman">AEP
    was
    incorporated under the laws of the State of New York in 1906 and reorganized
    in
    1925. It is a public utility holding company that owns, directly or indirectly,
    all of the outstanding common stock of its public utility subsidiaries and
    varying percentages of other subsidiaries.</font></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><br></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 18pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY: Times New Roman">The
    service areas of AEP&#8217;s public utility subsidiaries cover portions of the states
    of Arkansas, Indiana, Kentucky, Louisiana, Michigan, Ohio, Oklahoma, Tennessee,
    Texas, Virginia and West Virginia. The generating and transmission facilities
    of
    AEP&#8217;s public utility subsidiaries are interconnected and their operations are
    coordinated. Transmission networks are interconnected with extensive
    distribution facilities in the territories served. The public utility
    subsidiaries of AEP have traditionally provided electric service, consisting
    of
    generation, transmission and distribution, on an integrated basis to their
    retail customers. Restructuring legislation in Michigan, Ohio, Texas and
    Virginia has caused AEP public utility subsidiaries in those states to unbundle
    previously integrated regulated rates for their retail customers.</font></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><br></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 18pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY: Times New Roman">The
    AEP
    System is an integrated electric utility system and, as a result, the member
    companies of the AEP System have contractual, financial and other business
    relationships with the other member companies, such as participation in the
    AEP
    System savings and retirement plans and tax returns, sales of electricity
```

and
transportation and handling of fuel. The member companies of the AEP System
also
obtain certain accounting, administrative, information systems, engineering,
financial, legal, maintenance and other services at cost from a common provider,
AEPSC.

At
December 31, 2005, the subsidiaries of AEP had a total of 19,630 employees.
Because it is a holding company rather than an operating company, AEP has
no
employees. The public utility subsidiaries of AEP are:

APCo (organized
in Virginia in 1926) is engaged in the generation, transmission and distribution
of electric power to approximately 942,000 retail customers in the southwestern
portion of Virginia and southern West Virginia, and in supplying and marketing
electric power at wholesale to other electric utility companies, municipalities
and other market participants. At December 31, 2005, APCo and its whoally
owned
subsidiaries had 2,408 employees. Among the principal industries served by
APCo
are coal mining, primary metals, chemicals and textile mill products. In
addition to its AEP System interconnections, APCo also is interconnected
with
the following unaffiliated utility companies: Carolina Power & Light
Company, Duke Energy Corporation and Virginia Electric and Power Company.
APCo
has several points of interconnection with TVA and has entered into agreements
with TVA under which APCo and TVA interchange and transfer electric power
over
portions of their respective systems. APCo is a member of PJM.

CSPCo (organized
in Ohio in 1937, the earliest direct predecessor company having been organized
in 1883) is engaged in the generation, transmission and distribution of electric
power to approximately 710,000 retail customers in Ohio, and in supplying
and
marketing electric power at wholesale to other electric utilities,
municipalities and other market participants. At December 31, 2005, CSPCo
had
1,178 employeees. CSPCo’s service area is comprised of two areas in Ohio, which
include portions of twenty-five counties. One area includes the City of Columbus
and the other is a predominantly rural area in south central Ohio. Among
the
principal industries served are food processing, chemicals, primary metals,
electronic machinery and paper products. In addition to its AEP System
interconnections, CSPCo also is interconnected with the following unaffiliated
utility companies: CG&E, DP&L and Ohio Edison Company. CSPCo is a member
of PJM. Pursuant to an acquisition that closed on December 31, 2005, CSPCo
purchased the electric utility operations of Monongahela Power Company in

Ohio.
As a result, in January 2006 approximately 29,000 customers in six southeastern Ohio counties, together with the transmission and distribution used to serve such customers, were added to CSPCo’s service territory.

I&M (organized
in Indiana in 1925) is engaged in the generation, transmission and distribution of electric power to approximately 581,000 retail customers in northern and eastern Indiana and southwestern Michigan, and in supplying and marketing electric power at wholesale to other electric utility companies, rural electric cooperatives, municipalities and other market participants. At December 31, 2005, I&M had 2,633 employees. Among the principal industries served is primary metals, transportation equipment, electrical and electronic machinery, fabricated metal products, rubber and miscellaneous plastic products and chemicals and allied products. Since 1975, I&M has leased and operated the assets of the municipal system of the City of Fort Wayne, Indiana. In addition to its AEP System interconnections, I&M also is interconnected with the following unaffiliated utility companies: Central Illinois Public Service Company, CG&E, Commonwealth Edison Company, Consumers Energy Company, Illinois Power Company, Indianapolis Power & Light Company, Louisville Gas and Electric Company, Northern Indiana Public Service Company, PSI Energy Inc.
and Richmond Power & Light Company. I&M is a member of PJM.

KPCo (organized
in Kentucky in 1919) is engaged in the generation, transmission and distribution of electric power to approximately 176,000 retail customers in an area in eastern Kentucky, and in supplying and marketing electric power at wholesale to
other electric utility companies, municipalities and other market participants. At December 31, 2005, KPCo had 454 employees. In addition to its AEP System interconnections, KPCo also is interconnected with the following unaffiliated utility companies: Kentucky Utilities Company and East Kentucky Power Cooperative Inc. KPCo is also interconnected with TVA. KPCo is a member of PJM.

Kingsport
Power Company (organized
in Virginia in 1917) provides electric service to approximately 46,000 retail customers in Kingsport and eight neighboring communities in northeastern Tennessee. Kingsport Power Company does not own any generating facilities and is
a member of PJM. It purchases electric power from APCo for distribution to its
customers. At December 31, 2005, Kingsport Power Company had 55 employees.

OPCo (organized
in Ohio in 1907 and re-incorporated in 1924) is engaged in the generation, transmission and distribution of electric power to approximately 710,000 retail
customers in the northwestern, east central, eastern and southern sections of
Ohio, and in supplying and marketing electric power at wholesale to other electric utility companies, municipalities and other market participants. At
December 31, 2005, OPCo had 2,220 employees. Among the principal industries served by OPCo are primary metals, rubber and plastic products, stone, clay, glass and concrete products, petroleum refining and chemicals. In addition to
its AEP System interconnections, OPCo also is interconnected with the following unaffiliated utility companies: CG&E, The Cleveland Electric Illuminating Company, DP&L, Duquesne Light Company, Kentucky Utilities Company, Monongahela Power Company, Ohio Edison Company, The Toledo Edison Company and
West Penn Power Company. OPCo is a member of PJM.

PSO (organized
in Oklahoma in 1913) is engaged in the generation, transmission and distribution of electric power to approximately 514,000 retail customers in eastern and southwestern Oklahoma, and in supplying and marketing electric power at wholesale to other electric utility companies, municipalities, rural electric cooperatives and other market participants. At December 31, 2005, PSO had 1,176
employees. Among the principal industries served by PSO are natural gas and oil
production, oil refining, steel processing, aircraft maintenance, paper manufacturing and timber products, glass, chemicals, cement, plastics, aerospace manufacturing, telecommunications, and rubber goods. In addition to its AEP System interconnections, PSO also is interconnected with Ameren Corporation, Empire District Electric Co., Oklahoma Gas & Electric Co., Southwestern Public Service Co. and Westar Energy Inc. PSO is a member of SPP.

SWEPCo (organized
in Delaware in 1912) is engaged in the generation, transmission and distribution of electric power to approximately 450,000 retail customers in northeastern Texas, northwestern Louisiana and western Arkansas, and in supplying and marketing electric power at wholesale to other electric utility companies, municipalities, rural electric cooperatives and other market participants. At
December 31, 2005, SWEPCo had 1,498 employees. Among the principal industries served by SWEPCo are natural gas and oil production, petroleum refining, manufacturing of pulp and paper, chemicals, food processing, and metal refining. The territory served by SWEPCo also includes several military installations, colleges, and universities. In addition to its AEP System interconnections, SWEPCo is also interconnected with CLECO Corp., Empire District Electric

Co.,
Entergy Corp. and Oklahoma Gas & Electric Co. SWEPCo is a member of
SPP.

TCC
(organized
in Texas in 1945) is engaged in the generation (to an extremely limited extent),
transmission and sale of power to affiliated and non-affiliated entities
and the
distribution of electric power to approximately 729,000 retail customers
through
REPs in southern Texas, and (to a limited extent) in supplying and marketing
electric power at wholesale to other electric utility companies and market
participants. Under the Texas Act, TCC is completing the final stage of exiting
the generation business and has already sold most of its generation assets,
including STP. At December 31, 2005, TCC had 1,160 employees. Among the
principal industries served by TCC are oil and gas extraction, food processing,
apparel, metal refining, chemical and petroleum refining, plastics, and
machinery equipment. In addition to its AEP System interconnections, TCC
is a
member of ERCOT.

TNC (organized
in Texas in 1927) is engaged in the generation, transmission and sale of
power
to affiliated and non-affiliated entities and the distribution of electric
power
to approximately 189,000 retail customers through REPs in west and central
Texas, and in supplying and marketing electric power at wholesale to other
electric utility companies, municipalities, rural electric cooperatives and
other market participants. At December 31, 2005, TNC had 387 employees. Among
the principal industries served by TNC are agriculture and the manufacturing
or
processing of cotton seed products, oil products, precision and consumer
metal
products, meat products and gypsum products. The territory served by TNC
also
includes several military installations and correctional facilities. In addition
to its AEP System interconnections, TNC is a member of ERCOT.

WPCo
(organized
in West Virginia in 1883 and reincorporated in 1911) provides electric service
to approximately 41,000 retail customers in northern West Virginia. WPCo
does
not own any generating facilities. WPCo is a member of PJM. It purchases
electric power from OPCo for distribution to its customers. At December 31,
2005, WPCo had 59 employees.


```
  <div style="DISPLAY: block; MARGIN-LEFT: 18pt; TEXT-INDENT: 18pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT:
0pt" align="justify"><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY: Times New
Roman"><strong><em>AEGCo</em></strong></font><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY:
Times New Roman"><em> </em></font><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY: Times
New Roman">(organized
    in Ohio in 1982) is an electric generating company. AEGCo sells power at
    wholesale to I&amp;M and KPCo. AEGCo has no employees.</font></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><br></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 18pt; TEXT-INDENT: 18pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT:
0pt" align="justify"><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY: Times New
Roman"><strong><em>SERVICE
    COMPANY SUBSIDIARY</em></strong></font><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY:
Times New Roman"><strong> </strong></font></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 18pt; TEXT-INDENT: 18pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><br></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 18pt; TEXT-INDENT: 18pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY: Times New Roman">AEP
    also
    owns a service company subsidiary, AEPSC. AEPSC provides accounting,
    administrative, information systems, engineering, financial, legal, maintenance
    and other services at cost to the AEP System companies. The executive officers
    of AEP and certain of its public utility subsidiaries are employees of AEPSC.
    At
    December 31, 2005, AEPSC had 5,760 employees.</font></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><br></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 36pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT:
0pt" align="justify"><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY: Times New
Roman"><strong><em>CLASSES
    OF SERVICE</em></strong></font></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 36pt; TEXT-INDENT: 0pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><br></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 18pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"><font style="DISPLAY: inline; FONT-SIZE: 10pt; FONT-FAMILY: Times New Roman">The
    principal classes of service from which the public utility subsidiaries of
    AEP
    derive revenues and the amount of such revenues during the year ended December
    31, 2005 are as follows:</font></div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 18pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"> </div>
  <div style="DISPLAY: block; MARGIN-LEFT: 0pt; TEXT-INDENT: 18pt; LINE-HEIGHT: 1.25; MARGIN-RIGHT: 0pt"
align="justify"> </div>
</div>
```


cleaned form

PART I ITEM 1. BUSINESS GENERAL OVERVIEW AND DESCRIPTION OF SUBSIDIARIES AEP was incorporated under the laws of the State of New York in 1906 and reorganized in 1925. It is a public utility holding company that owns, directly or indirectly all of the outstanding common stock of its public utility subsidiaries and varying percentages of other subsidiaries. The service areas of AEPs public utility subsidiaries cover portions of the states of Arkansas, Indiana, Kentucky, Louisiana, Michigan, Ohio, Oklahoma, Tennessee, Texas, Virginia and West Virginia. The generating and transmission facilities of AEPs public utility subsidiaries are interconnected and their operations are coordinated. Transmission networks are interconnected with extensive distribution facilities in the territories served. The public utility subsidiaries of AEP have traditionally provided electric service, consisting of generation, transmission and , on an integrated basis to their retail customers. Restructuring legislation in Michigan, Ohio, Texas and Virginia has caused AEP public utility subsidiaries in those states to unbundle previously integrated regulated rates for their retail customers. The AEP System is an integrated electric utility system and, as a result, the member companies of the AEP System have contractual, financial and other business relationships with the other member companies, such as participation in the AEP System savings and retirement plans and tax returns, sales of electricity and transportation and handling of fuel. The member companies of the AEP System also obtain certain accounting, administrative, information systems,

engineering, financial, legal, maintenance and other services at cost from a common provider, AEPSC. At December 31, 2005, the subsidiaries of AEP had a total of 19,630 employees. Because it is a holding company rather than an operating company, AEP has no employees. The public utility subsidiaries of AEP are: APCo (organized in Virginia in 1926) is engaged in the generation, transmission and of electric power to approximately 942,000 retail customers in the southwestern portion of Virginia and southern West Virginia, and in supplying and marketing electric power at wholesale to other electric utility companies, municipalities and other market participants. At December 31, 2005, APCo and its whoally owned subsidiaries had 2,408 employees. Among the principal industries served by APCo are coal mining, primary metals, chemicals and textile mill products. In addition to its AEP System interconnections, APCo also is interconnected with the following unaffiliated utility companies: Carolina Power & Light Company, Duke Energy Corporation and Virginia Electric and Power Company. APCo has several points of interconnection with TVA and has entered into agreements with TVA under which APCo and TVA interchange and transfer electric power over portions of their respective systems. APCo is a member of PJM. CSPCo (organized in Ohio in 1937, the earliest direct predecessor company having been organized in 1883) is engaged in the generation, transmission and distribution of electric power to approximately 710,000 retail customers in Ohio, and in supplying and marketing electric power at wholesale to other electric utilities, municipalities and other market participants. At December 31, 2005, CSPCo had 1,178 employeees. CSPCos service area is comprised of two areas in Ohio, which include portions of twenty-five counties. One area includes the City of Columbus and the other is a predominantly rural area in south central Ohio. Among the principal industries served are food processing, chemicals, primary metals, electronic machinery and paper products. In addition to its AEP System interconnections, CSPCo also is interconnected with the following unaffiliated utility companies: CG&E, DP&L and Ohio Edison Company. CSPCo is a member of PJM. Pursuant to an acquisition that closed on December 31, 2005, CSPCo purchased the electric utility operations of Monongahela Power Company in Ohio. As a result, in January 2006 approximately 29,000 customers in six southeastern Ohio counties, together with the transmission and distribution used to serve such customers, were added to CSPCos service territory. I&M (organized in Indiana in 1925) is engaged in the generation, transmission and distribution of electric power to approximately 581,000 retail customers in northern and eastern Indiana and southwestern Michigan, and in supplying and marketing electric power at wholesale to other electric utility companies, rural electric cooperatives, municipalities and other market participants. At December 31, 2005, I&M had 2,633 employees. Among the principal industries served are primary metals, transportation equipment, electrical and electronic machinery, fabricated metal products, rubber and miscellaneous plastic products and chemicals and allied products. Since 1975, I&M has leased and operated the assets of the municipal system of the City of Fort Wayne, Indiana. In addition to its AEP System interconnections, I&M also is interconnected with the following unaffiliated utility companies: Central Illinois Public Service Company, CG&E, Commonwealth Edison Company, Consumers Energy Company, Illinois Power Company, Indianapolis Power & Light Company, Louisville Gas and Electric Company, Northern Indiana Public Service Company, PSI Energy Inc. and Richmond Power & Light Company. I&M is a member of PJM. KPCo (organized in Kentucky in 1919) is engaged in the generation, transmission and distribution of electric power to approximately 176,000 retail customers in an area in eastern Kentucky, and in supplying and marketing electric power at wholesale to other electric utility companies, municipalities and other market participants. At December 31, 2005, KPCo had 454 employees. In addition to its AEP System interconnections, KPCo also is interconnected with the following unaffiliated utility companies: Kentucky Utilities Company and East Kentucky Power Cooperative Inc. KPCo is also interconnected with TVA. KPCo is a member of PJM. Kingsport Power Company(organized in Virginia in 1917) provides electric service to approximately 46,000 retail customers in Kingsport and eight neighboring communities in northeastern Tennessee. Kingsport Power Company does not own any generating facilities and is a member of PJM. It purchases electric power from APCo for distribution to its customers. At December 31, 2005, Kingsport Power Company had 55 employees. OPCo (organized in Ohio in 1907 and re-incorporated in 1924) is engaged in the generation, transmission and distribution of electric power to approximately 710,000 retail customers in the northwestern, east central, eastern and southern sections of Ohio, and in supplying and marketing electric power at wholesale to other electric utility companies, municipalities and other market participants. At December 31, 2005, OPCo had 2,220 employees. Among the principal industries served by OPCo is primary metals, rubber and plastic products, stone, clay, glass and concrete products, petroleum refining and chemicals. In addition to its AEP System interconnections, OPCo also is interconnected with the following unaffiliated utility companies: CG&E, The Cleveland Electric Illuminating Company, DP&L, Duquesne Light Company, Kentucky Utilities Company, Monongahela Power Company, Ohio Edison Company, The Toledo Edison Company and West Penn Power Company. OPCo is a member of PJM. PSO (organized in Oklahoma in 1913) is engaged in the generation, transmission and distribution of electric power to approximately 514,000 retail customers in eastern and southwestern Oklahoma, and in supplying and marketing electric power at wholesale to other electric utility companies, municipalities, rural electric cooperatives and other market participants. At December 31, 2005, PSO had 1,176 employees. Among the principal industries served by PSO are natural gas and oil production, oil refining, steel processing, aircraft maintenance, paper manufacturing and timber products, glass, chemicals, cement, plastics, aerospace

manufacturing, telecommunications, and rubber goods. In addition to its AEP System interconnections, PSO also is interconnected with Ameren Corporation, Empire District Electric Co., Oklahoma Gas & Electric Co., Southwestern Public Service Co. and Westar Energy Inc. PSO is a member of SPP. SWEPCo

sentence errors
- **comma error** It is a public utility holding company that owns, directly or indirectly**,** all of the outstanding common stock of its public utility subsidiaries and varying percentages of other subsidiaries.
- **sp error** At December 31, 2005, APCo and its **whoally** owned subsidiaries had 2,408 employees.
- **sp error** At December 31, 2005, CSPCo had 1,178 **employeees**.
- **agreement error** Among the principal industries served by OPCo **is** primary metals, rubber and plastic products, stone, clay, glass and concrete products, petroleum refining and chemicals.

csv output
,,,,,,177,2,2