

Name Date Place

Extraction in unstructured text

Automatically scan machine-readable text
to locate name, date, and place
information

The Problem

It's difficult to:

- Find pertinent information in long documents
- Make accurate queries for unknown entities
- Make queries that compensate for all variations
 - (spelling, alternate names, format)

Our Proposal

Create a tool that will find all the
locations of names, dates,
and places within a document.

Examples of Family Stories - Windows Internet Explorer

http://www.familyhistoryproducts.com/examples-of-family-stories.html

File Edit View Favorites Tools Help

☆ Favorites Center

Chris Brown Music Profil... Gmail - Inbox (285) - dav... Examples of Family St... x

Home RSS Print Page Tools

NDP Extractor Search x

Tips and Ideas

Journal Prompts
Journaling Tips
Oral History
Oral History Recording

Subscribe to The Family History Product Newsletter!

Email

Name

Then
Subscribe

Don't worry -- your e-mail address is totally secure. I promise to use it only to send you Family History Help & Product Review.

Subscribe To This Site

RSS

Google


MY Yahoo!

MY MSN

newsgator

Bloglines

A True Story About My Family ... I Couldn't Believe It!



A decade ago, I came across an individual with my same last name of Boyter. Boyter isn't that common of a name, as you know.

I was 30 years old and, up to that point in my life, I had never crossed paths with a person with my surname and not known immediately who they were and how we were related. **I was puzzled and intrigued!**


On a side note, **this man lived only two hours north of where I grew up and my grandparents didn't know who he was either.**

We sat and talked for several hours that day, swapping family information, stories and looking at family photographs. As we talked I couldn't help but notice how his general look and mannerisms reminded me of my grandpa Boyter.

He told a story of his ancestor James Boyter who, in the 1870s, traversed the Atlantic Ocean from Scotland with his older brother Alexander.

According to family "legend" one or both of the brothers had gotten into trouble for hunting on royal property. The story went on that these two, in an attempt to escape the law, fled to the United States.


Record Your Lifestory to CD



Keeping a [diary](#) or [journal](#) has never been easier or more convenient! The Journal automatically brings up a new entry for you every day, and provides a calendar for easy review.

[DOWNLOAD The JOURNAL 4.0](#)

Things to Do With Your Family Photos





Internet | Protected Mode: On

Tuesday, March 10, 2009

Examples of Famil... NDPextractor NDPpresentation... NamesDatesPlace... Macromedia Fire... 4:53 PM

Examples of Family Stories - Windows Internet Explorer

http://www.familyhistoryproducts.com/examples-of-family-stories.html

File Edit View Favorites Tools Help

☆

★

Favorites Center

Chris Brown Music Profil...

Gmail - Inbox (285) - dav...

Examples of Family St...

Home

Print

Page

Tools

NDP Extractor

Search

Jump to: [Names](#), [Dates](#), [Places](#)

Names

Boyer12345

James12

Alexander123

Robert1

Stevens1

Dates

1800

1800s1

1870

18701

2000-20091

2007-20091

Places

Scotland12

United States12

U.S.12

Utah12

Tips and Ideas

Journal Prompts

Journaling Tips

Oral History

Oral History Recording

Subscribe to The Family History Product Newsletter!

Email

Name

Then

Subscribe

Don't worry -- your e-mail address is totally secure.

I promise to use it only to send you Family History Help & Product Review.

Subscribe To This Site

RSS

Google


MY Yahoo!

MY MSN

newsgator

Bloglines

A True Story About My Family ... I Couldn't Believe It!



A decade ago, I came across an individual with my same last name of Boyter. Boyter isn't that common of a name, as you know.

I was 30 years old and, up to that point in my life, I had never crossed paths with a person with my surname and not known immediately who they were and how we were related. **I was puzzled and intrigued!**


On a side note, **this man lived only two hours north of where I grew up and my grandparents didn't know who he was either.**

We sat and talked for several hours that day, swapping family information, stories and looking at family photographs. As we talked I couldn't help but notice how his general look and mannerisms reminded me of my grandpa Boyter.

He told a story of his ancestor James Boyter who, in the 1870s, traversed the Atlantic Ocean from Scotland with his older brother Alexander.

According to family "legend" one or both of the brothers had gotten into trouble for hunting on royal property. The story went on that these two, in an attempt to escape the law, fled to the United States.


Record Your Lifestory to CD



Keeping a [diary](#) or [journal](#) has never been easier or more convenient! The Journal automatically brings up a new entry for you every day, and provides a calendar for easy review.

[DOWNLOAD The JOURNAL 4.0](#)

Things to Do With Your Family Photos



smugmug

The ultimate in

Tuesday, March 10, 2009

4:53 PM

Examples of Family Stories - Windows Internet Explorer

http://www.familyhistoryproducts.com/examples-of-family-stories.html

File Edit View Favorites Tools Help

★

★

Favorites Center

Chris Brown Music Profil...

Gmail - Inbox (285) - dav...

Examples of Family St...

Home

Print

Page

Tools

NDP Extractor

Search

Jump to: [Names](#), [Dates](#), [Places](#)

Names

Boyer12345

James12

Alexander123

Robert1

Stevens1

Dates

1800

1800s1

1870

18701

2000-20091

2007-20091

Places

Scotland12

United States12

U.S.12

Utah12

Tips and Ideas

Journal Prompts

Journaling Tips

Oral History

Oral History Recording

Subscribe to The Family History Product Newsletter!

Email

Name

Then

Subscribe

Don't worry -- your e-mail address is totally secure.

I promise to use it only to send you Family History Help & Product Review.

Subscribe To This Site

RSS

Google


MY Yahoo!

MY MSN

newsgator

Bloglines

A True Story About My Family ... I Couldn't Believe It!



A decade ago, I came across an individual with my same last name of Boyter. Boyter isn't that common of a name, as you know.

I was 30 years old and, up to that point in my life, I had never crossed paths with a person with my surname and not known immediately who they were and how we were related. **I was puzzled and intrigued!**


On a side note, **this man lived only two hours north of where I grew up and my grandparents didn't know who he was either.**

We sat and talked for several hours that day, swapping family information, stories and looking at family photographs. As we talked I couldn't help but notice how his general look and mannerisms reminded me of my grandpa Boyter.

He told a story of his ancestor James Boyter who, in the 1870s, traversed the Atlantic Ocean from Scotland with his older brother Alexander.

According to family "legend" one or both of the brothers had gotten into trouble for hunting on royal property. The story went on that these two, in an attempt to escape the law, fled to the United States.


Record Your Lifestory to CD




Keeping a [diary](#) or [journal](#) has never been easier or more convenient! The Journal automatically brings up a new entry for you every day, and provides a calendar for easy review.

[DOWNLOAD The JOURNAL 4.0](#)

Things to Do With Your Family Photos





Internet | Protected Mode: On

Tuesday, March 10, 2009

4:53 PM

How we plan to do it

Four step Algorithm

1. Convert the content to plain text.
2. Convert the text from a sequence of characters to a sequence of categorized tokens.
3. Identify the complete names, dates, and places with a lexical analyzer. (combine tokens)
4. Format the results.

Convert to plain text

<p class="MsoPlainText" style="line-height:150%;">Cities on a Saturday are often such interesting places: full of people, full of cars, full of the hustle and bustle of modern life. And Leicester is no exception. I was born there so I can speak from personal experience. But something was different last Saturday. There were more people, more cars and much more hustle and bustle than I had ever seen or heard before. </p>

Cities on a Saturday are often such interesting places: full of people, full of cars, full of the hustle and bustle of modern life. And Leicester is no exception. I was born there so I can speak from personal experience. But something was different last Saturday. There were more people, more cars and much more hustle and bustle than I had ever seen or heard before.

<p class="MsoPlainText" style="line-height:150%;"> I�d gone into town with my mates that Saturday - as we always do. We caught the same No. 149 bus from Oadby � that�s a small town south of Leicester. Nothing unusual in that. The journey was as predictable as ever � I�m so used to it. I can�t even remember getting on the bus; but I can certainly remember getting off�

Id gone into town with my mates that Saturday - as we always do. We caught the same No. 149 bus from Oadby thats a small town south of Leicester. Nothing unusual in that. The journey was as predictable as ever Im so used to it. I cant even remember getting on the bus; but I can certainly remember getting off

Tokenize and Categorize

- **Divide the text into organizable pieces**
 - Tokenize the input on white space and punctuation
- **Identify strings of characters as simple tokens classified as parts of names, dates, or places**
 - Use a Name Authority to determine parts of names
 - Use a Place Authority to determine parts of places
 - Use research done by Robert Lyon to identify dates

Lexically analyze

Place	→ <i>city</i> PlacesFollowingCity CountyStateCountry StateCountry <i>country</i>
PlacesFollowingCity	→ PossibleComma CountyOrStateOrCountry ε
CountyOrStateOrCountry	→ CountyStateCountry StateCountry <i>country</i>
CountyStateCountry	→ <i>county</i> CountyDesignator PlacesFollowingCounty StateCountry <i>Country</i>
CountyDesignator	→ <i>Left_Paren</i> CountyOrParish <i>Right_Paren</i> CountyOrParish
CountyOrParish	→ <i>countyWord</i> <i>parishWord</i>
PlacesFollowingCounty	→ PossibleComma StateOrCountry ε
StateOrCountry	→ StateCountry <i>country</i>
StateCountry	→ <i>state</i> StateDesignator PlacesFollowingState
StateDesignator	→ <i>Left_Paren</i> <i>stateWord</i> <i>Right_Paren</i> <i>stateWord</i>
PlacesFollowingState	→ PossibleComma <i>country</i>
PossibleComma	→ <i>Comma</i> ε

Name	→ <i>title</i> SimpleName SimpleName
SimpleName	→ NameWithGivenNames LastName
NameWithGivenNames	→ GivenNames LastNameAndSuffix
LastNameAndSuffix	→ LastName GenerationId ε
GivenNames	→ <i>givenName</i> OptionalPeriod RemainingGivenNames
RemainingGivenNames	→ GivenNames ε
LastName	→ <i>surname</i> RemainingSurnames
RemainingSurnames	→ <i>dash</i> <i>surname</i> SurnameList
SurnameList	→ <i>surname</i> SurnameList ε
GenerationId	→ <i>suffix</i> ε
OptionalPeriod	→ <i>period</i> ε

Create completed name, date, and place results by combining our categorized tokens using these regular grammars

Date Identification

September 1, 1997	- Original
1 September 1997	- Alternative ordering
Sept. 1, 1997	- Month abbreviation
Sept 1, 1997	- Alternate punctuation
Sept 1, '97	- Year abbreviation
Sept 1	- Assumed year
September 1997	- No day of the month
09/01/1997	- Numeric format
September 1st 1997	- Ordinal day of the month
1st of September 1997	- Internal preposition
after Sept 1, 1997	- Altering preposition

[Lyon2000] Lyon, Robert W., *Identification of temporal phrases in natural language*, Masters Thesis, Brigham Young University. Dept. of Computer Science, 2000

Format results

Jump to: [Names](#), [Dates](#), [Places](#)

Names

Boyter	1	2	3	4	5
James	1	2			
Alexander	1	2	3		
Robert	1				
Stevens	1				

Dates

1800	
1800s	1
1870	
1870	1
2000-2009	1
2007-2009	1

Places

Scotland	1	2
United States	1	2
U.S.	1	2
Utah	1	2





Time line

- Summer '09
 - Recruit BYU CS students for capstone
 - Further research and design of the project
 - Find/Develop solutions for name and place authority requirements
- Fall Semester '09
 - Implement CS598R capstone project to develop the NDPextractor
- December '09
 - Finish CS598R capstone project

Questions?

