How we tripled our encoding speed in the Digital Victorian Periodical Project

Kaitlyn Fralick, Kailey Fukushima, Martin Holmes, and Sarah Karlson
(University of Victoria)

Digital Victorian Periodical Poetry

Dr. Alison Chapman (PI), Martin Holmes, & team
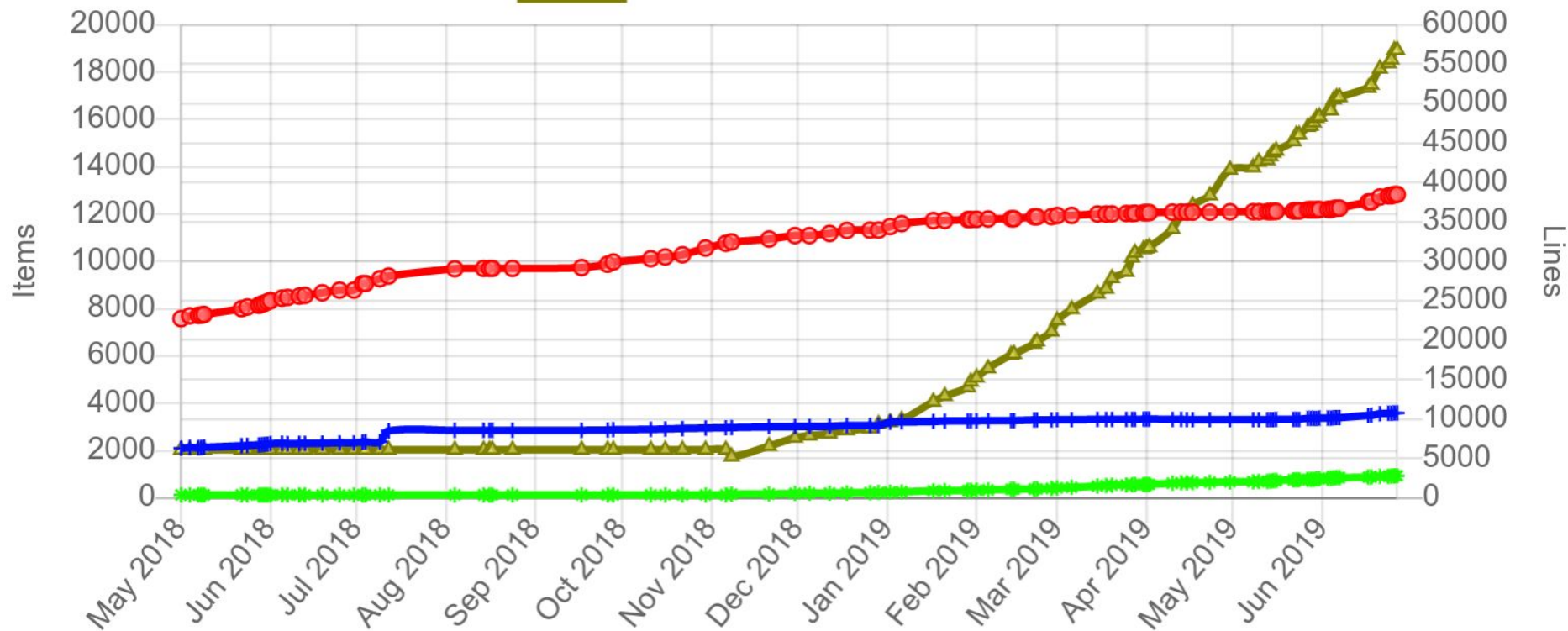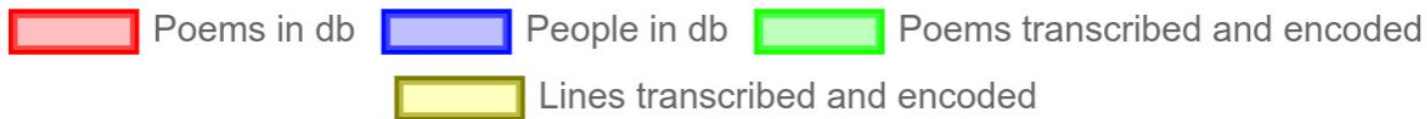@poetry_digital

The Team

Kaitlyn Fralick

Kailey Fukushima

Sarah Karlson

The problem:

- About 2,000 poems to encode

- Limited time and money

- Projections showing we were not going to make our target

- Encoders frustrated by repetitive and tedious tasks

SONG.

BY THE LADY E. STUART WORTLEY.

Awake to the sound of my sighing,
  Breeze, gentle breeze of the night!
Like an echo awake thou, replying
  To my grief's bitter wailing aright.

Let me think it is Nature thus mourning,
  Deep Nature and sorrowful Night;
For my soul and my heart I am scorning,
  Thus bowed beneath Sorrow's cold blight.

Wake then to the sound of my sighing—
  Breeze! quiet breeze of the night!
In the tones of thy sinking and dying,
  Perchance thou may'st chime it aright.

And oh! not alone with my sighing,
  Be burthened, sweet breeze of the night;
This soul—to which none are replying,
  Bear *that* too with thee on thy flight!

Eisgrub.

1. <u>Automation of teiHeader and OCR</u>

- Automatic TEI header (SQL database → XML → TEI)

- Automatic OCR stores output as a `<!--comment-->` in the body of the TEI file

- Encoder proofs and corrects OCR output

```xml
<teiHeader>
    <fileDesc>
        <titleStmt>
            <title>Song</title>
            <!--Do not edit author information; this is maintained in the
    database and generated automatically.-->
            <author ref="dvpp:prs_3331">
                <persName>
                    <forename>Emmeline Charlotte Elizabeth</forename>
                    <surname>Stuart-Wortley</surname>
                    <name type="displayName">Stuart-Wortley, Emmeline</name>
                </persName>
            </author>
        </titleStmt>
        <publicationStmt>
            <publisher>University of Victoria Digital Victorian Periodical Poetry
                Project</publisher>
            <pubPlace>Victoria, BC, Canada</pubPlace>
            <availability>
                <p>In the public domain</p>
            </availability>
            <!-- Do not edit this date; it's generated automatically. -->
            <date notAfter="2019" notBefore="2016"/>
        </publicationStmt>
        <notesStmt>
            <!-- Do not edit the note elements below.
     They are automatically generated from the database. -->
            <note>Poem signed "The Lady E. Stuart Wortley" and ends with "Eisgrub." (AC)</note>
        </notesStmt>
        <sourceDesc>
            <!-- Do not edit the contents of the sourceDesc/bibl element.
     It is generated from the database, and any errors
     must be fixed there. -->
            <bibl corresp="dvpp:bib_18">
                <title level="j">The Keepsake</title>
                <biblScope unit="volume">1840</biblScope>
                <biblScope unit="page" from="191" to="191">191</biblScope>
                <date when="1840"/>
            </bibl>
```

```xml
    <text>
        <body>
            <div type="poem">

<!-- OUTPUT OF OCR PROCESS:

SONG.
BY THE LADY E. STUART WORTLEY.

Awake to the sound of my sighing,
Breeze, gentle breeze of the night !
Like an echo awake thou, replying
To my grief's bitter wailing aright.

Let me think it is Nature thus mourning,
Deep Nature and sorrowful Night ;
For my soul and my heart I am scorning,
Thus bowed beneath Sorrow's cold blight.

Wake then to the sound of my sighing—
Breeze! quiet breeze of the night!
In the tones of thy sinking and dying,
Perchance thou may'st chime it aright.

And oh! not alone with my sighing,
Be burthened, sweet breeze of the night ;
This soul—to which none are replying,
Bear that too with thee on thy flight !

Eisgrub.
-->

            </div>
        </body>
    </text>
```

2. <u>Schematron Quickfix for \<lg> and \<l></u>

- Encoder removes corrected OCR from \<!--comment-->

- Encoder uses <u>Schematron Quickfix</u> (by right-clicking on \<div>) to autotag all lines \<l> and line groups \<lg> in the text

```
<div type="poem">
```

Awake to

Breeze, gentle breeze of the night !
Like an echo awake thou, replying
To my grief's bitter wailing aright.

Let me think it is Nature thus mourning,
Deep Nature and sorrowful Night ;
For my soul and my heart I am scorning,
Thus bowed beneath Sorrow's cold blight.

Wake then to the sound of my sighing—
Breeze! quiet breeze of the night!
In the tones of thy sinking and dying,
Perchance thou may'st chime it aright.

And oh! not alone with my sighing,
Be burthened, sweet breeze of the night ;
This soul—to which none are replying,
Bear that too with thee on thy flight !

```
<div type="poem">
```

> ⊘ **Untagged text should probably be tagged as a stanza or a paragraph.**
> ↻ Tag this block as one or more stanzas.

```
Awake to
Breeze, gentle breeze of the night !
Like an echo awake thou, replying
To my grief's bitter wailing aright.
```

```xml
<div type="poem">
    <lg rhyme="PENDING">
        <l>Awake to the sound of my sighing, </l>
        <l>Breeze, gentle breeze of the night ! </l>
        <l>Like an echo awake thou, replying </l>
        <l>To my grief's bitter wailing aright.</l>
    </lg>

    <lg rhyme="PENDING">
        <l>Let me think it is Nature thus mourning, </l>
        <l>Deep Nature and sorrowful Night ; </l>
        <l>For my soul and my heart I am scorning,</l>
        <l>Thus bowed beneath Sorrow's cold blight.</l>
    </lg>

    <lg rhyme="PENDING">
        <l>Wake then to the sound of my sighing— </l>
        <l>Breeze! quiet breeze of the night! </l>
        <l>In the tones of thy sinking and dying, </l>
        <l>Perchance thou may'st chime it aright.</l>
    </lg>

    <lg rhyme="PENDING">
        <l>And oh! not alone with my sighing, </l>
        <l>Be burthened, sweet breeze of the night ; </l>
        <l>This soul—to which none are replying, </l>
        <l>Bear that too with thee on thy flight !</l>
    </lg>
```

```
<body>
    <div type="poem">

        <head>SONG.</head>
        <head>BY THE LADY E. STUART WORTLEY.</head>

        <lg rhyme="PENDING">                    [5 lines]

        <lg rhyme="PENDING">                    [5 lines]

        <lg rhyme="PENDING">                    [5 lines]

        <lg rhyme="PENDING">                    [5 lines]

        <byline><placeName>Eisgrub</placeName>.</byline>

    </div>
</body>
```
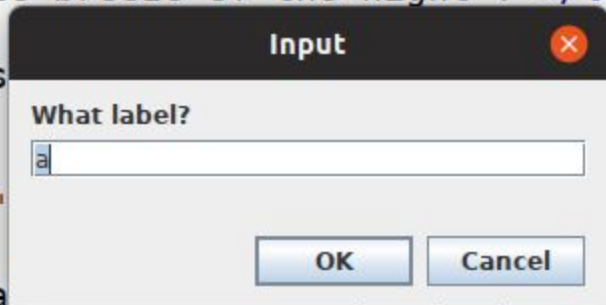
3. Keystroke shortcut for rhyme

- [Control + Alt + 0] wraps selected text in a <rhyme> tag, prompts the encoder to add a rhyme @label

- This keystroke shortcut employs an XPath regular expression to identify a likely rhyme @type (masculine/feminine) based on syllable count

```
<lg rhyme="abab">
    <l>Awake to the sound of my sighing, </l>
    <l>Breeze, gentle breeze of the night ! </l>
    <l>Like an echo
    <l>To my grief's
</lg>

<lg rhyme="PENDING"
    <l>Let me think                    </l>
    <l>Deep Nature a
    <l>For my soul and my heart I am scorning </l>
```

Input   ✕

What label?

```
a
```

[OK]   [Cancel]

```
<rhyme type="dvpp:rhymeFeminine" label="a">sighing</rhyme>, </l>
the night ! </l>
```
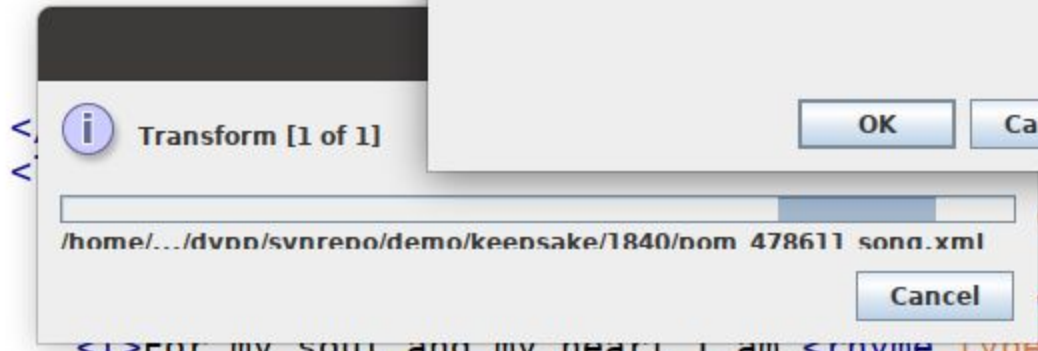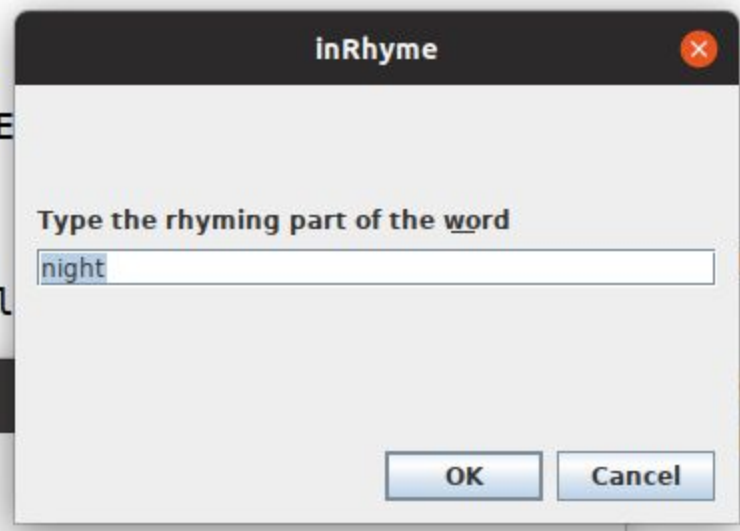
4. XSLT transformation for rhyme

- Our project reuses rhyme labels. If "love" is labelled "a" in stanza 1, "dove" will be labelled "a" in stanza 57.

- An XSLT transformation scenario helps the encoder to find repeated rhyme labels within long poems.

- It does this by searching the entire corpus of encoded rhymes for likely matches.

```
div type="poem">

   <head>SONG.</head>
   <head>BY THE LADY E

   <lg rhyme="abab">
      <l>Awake to the                                          meFe
      <l>Breeze, gentl                                         :rhy
```

inRhyme   ✕

Type the rhyming part of the word

night

                                                    OK   Cancel

ⓘ **Transform [1 of 1]**

/home/.../dvpp/svnrepo/demo/keepsake/1840/pom_478611_song.xml

                                                                    Cancel
```
   <l>For my sout and my heart i am <rhyme type="dvpp:rh
```

```
POTENTIAL RHYMES FOR night FOUND IN POEM 478611

night  (label = b; pom_478611_song)
a | right  (label = b; pom_478611_song)
Night  (label = b; pom_478611_song)
blight  (label = b; pom_478611_song)
flight  (label = b; pom_478611_song)
```

5. CSS efficiency and accuracy

- Encoder uses <rendition> (with @selector) in the encodingDesc of the teiHeader to describe the overall layout of poem.

- Encoder uses an XSLT transformation scenario to get instant visual feedback on TEI and CSS encoding and alters their encoding as needed.

SONG.

BY THE LADY E. STUART WORTLEY.

AWAKE to the sound of my sighing,
    Breeze, gentle breeze of the night!
Like an echo awake thou, replying
    To my grief's bitter wailing aright.

Let me think it is Nature thus mourning,
    Deep Nature and sorrowful Night;
For my soul and my heart I am scorning,
    Thus bowed beneath Sorrow's cold blight.

Wake then to the sound of my sighing—
    Breeze! quiet breeze of the night!
In the tones of thy sinking and dying,
    Perchance thou may'st chime it aright.

And oh! not alone with my sighing,
    Be burthened, sweet breeze of the night;
This soul—to which none are replying,
    Bear *that* too with thee on thy flight!

EISGRUB.

sisted of bare walls, a bench, and some dried leaves, on which
lay the body of the long lost Francesco; and that there was,
separated from this part by a partition, a shrine and tomb
of exquisite workmanship, strewn with flowers hardly withered,
and on which no record was placed but the word

BEATRICE!

B. H.

SONG.

BY THE LADY E. STUART WORTLEY.

AWAKE to the sound of my sighing,
  Breeze, gentle breeze of the night!
Like an echo awake thou, replying
  To my grief's bitter wailing aright.

Let me think it is Nature thus mourning,
  Deep Nature and sorrowful Night;
For my soul and my heart I am scorning,
  Thus bowed beneath Sorrow's cold blight.

Wake then to the sound of my sighing—
  Breeze! quiet breeze of the night!
In the tones of thy sinking and dying,
  Perchance thou may'st chime it aright.

And oh! not alone with my sighing,
  Be burthened, sweet breeze of the night;
This soul—to which none are replying,
  Bear *that* too with thee on thy flight!

EISGRUB.

```
<encodingDesc>
   <tagsDecl>
      <rendition selector="lg">margin-left: 6em;</rendition>
      <rendition selector="l:nth-of-type(even)">margin-left: 1em;</rendition>
   </tagsDecl>
</encodingDesc>
```

6. <u>Feedback and error reporting</u>

- In addition to providing visual feedback for the encoder's CSS, the <u>HTML rendering</u> reports on errors such as:

  - Inconsistency between the @rhyme pattern in the \<lg\> and the sequence of \<rhyme\> @labels within the stanza

  - Omission of a @rhyme pattern in the \<lg\>

  - Omission of a declared width in the poem \<div\>

SONG.

BY THE LADY E. STUART WORTLEY.

Awake to the sound of my sighing,　　　　1
　　Breeze, gentle breeze of the night !　　2
Like an echo awake thou, replying　　　　3
　　To my grief's bitter wailing aright.　　4

Let me think it is Nature thus mourning,　　5
　　Deep Nature and sorrowful Night ;　　6
For my soul and my heart I am scorning,　　7
　　Thus bowed beneath Sorrow's cold blight.　　8

Wake then to the sound of my sighing—　　9
　　Breeze! quiet breeze of the night!　　10
In the tones of thy sinking and dying,　　11
　　Perchance thou may'st chime it aright.　　12

And oh! not alone with my sighing,　　　13
　　Be burthened, sweet breeze of the night ;　　14
This soul—to which none are replying,　　15
　　Bear that too with thee on thy flight !　　16

Eisgrub.

7. Sonic devices

- The <u>HTML rendering</u> uses a simple similarity metric to notify the encoder of potential <u>sonic devices</u> within the poem (e.g. choruses, refrains, anaphoras)

- Encoder evaluates the sonic device report and tags confirmed sonic devices within the text

SONG.

BY THE LADY E. STUART WORTLEY.

Awake to the sound of my sighing,   1
   Breeze, gentle breeze of the night !   2
Like an echo awake thou, replying   3
   To my grief's bitter wailing aright.   4

Let me think it is Nature thus mourning,   5
   Deep Nature and sorrowful Night ;   6
For my soul and my heart I am scorning,   7
   Thus bowed beneath Sorrow's cold blight.   8

Wake then to the sound of my sighing—   9
   Breeze! quiet breeze of the night!   10
In the tones of thy sinking and dying,   11
   Perchance thou may'st chime it aright.   12

And oh! not alone with my sighing,   13
   Be burthened, sweet breeze of the night ;   14
This soul—to which none are replying,   15
   Bear that too with thee on thy flight !   16

Eisgrub.

8. Tracking labour

- After a prompt from the DVPP <u>schematron</u>, encoder adds their <respStmt> (with a unique @xml:id) to teiHeader

31   </notesStmt>

Don't forget to insert <respStmt> elements to describe the roles played by team members in creating this file.

Text   **Grid**   **Author**

```
<teiHeader>
    <fileDesc>
        <titleStmt>
            <title           ⓘ Don't forget to insert <respStmt> elements to describe the roles played by team members in creat ...
            <!--Do              ⟲ Insert a respStmt
    database and generated automatically.-->
            <author ref="dvpp:prs_3331">
                <persName>
                    <forename>Emmeline Charlotte Elizabeth</forename>
```

```xml
<respStmt>
    <resp ref="resp:mrk">
        <date when="2019"/>
    </resp>
    <persName ref="dvpp:prs_ed_xxx">
        <forename>FORENAME</forename>
        <surname>SURNAME</surname>
    </persName>
</respStmt>
</titleStmt>
```

Summary

- Pre-build as much as possible.
- Automate repetitive tagging.
- Provide instant rendering with diagnostics.
- Make your programmer do encoding!

For code and additional details:
http://bit.ly/encoding_speed

@poetry_digital