

Welcome!

Session 8: Tuesday, 3:00 – 5:00 pm

- Discussion: review testing results
- Hands-on: update designs and/or document lessons from testing
- Let's talk about text!

Discussion: test results

- What did you hear?
- How can you translate that into updates, notes for future designs?
- What would your next test focus on?

Additional review and critique

- Can your task be made smaller?
- Is the 'call to action' clear and powerful enough to convert sightseers into participants?
- Where can you remove unnecessary 'friction' from the process?
- What might demotivate participants, and how can you avoid that?
- How are contributions validated for correctness?

Additional review and critique

- Audience
 - Which motivations for starting, continuing are you appealing to?
 - How are participants rewarded?
 - How will they know they're doing their tasks well?

Additional review and critique

- Content
 - Does the input content need processing or cleaning first?
 - What kinds of crowdsourced tasks create the output data?
 - How might you validate the results?
 - Can you design validation tasks into the project?

For more inspiration

- Go back to the list of projects listed for speed-dating
- Ask for suggestions for related projects

For more inspiration

- Go back to the list of projects listed for speed-dating
- Ask for suggestions for related projects

Let's talk about text

The problem with “scan-and-dump”

One Volunteer's Story

Nat Wooding

- Retired data analyst
- 100 pages of Julia Brumfield's diaries transcribed and indexed in six months
- No relation to diarist

One Volunteer's Story

Nat Wooding

- Retired data analyst
- 100 pages of Julia Brumfield's diaries transcribed and indexed in six months
- No relation to diarist
- Great-uncle was diarist's letter carrier, also named **Nat Wooding**

1810. Well hope I won't get
very worse that wooding
carried the mail today
I am very glad that he
was able to go - I got
a letter from him this

FromThePage

Julia Brumfield Diaries — 1920

page

transcribe

versions

[Previous Page](#)

Wednesday, February 25, 1920

A colde day. The ground was covered in snow this morning. The boys here did no try to work out.

Josie got dinner.

I sowed on my quilt most of the time.

I am not well. Hope I wont get any worse.

Nat Wooding carried the mail today. I am very glad that he was able to go. I got a

OCR Under the Hood

OCR Under the Hood



Cerceris compacta

OCR Under the Hood

OCR Text Output:

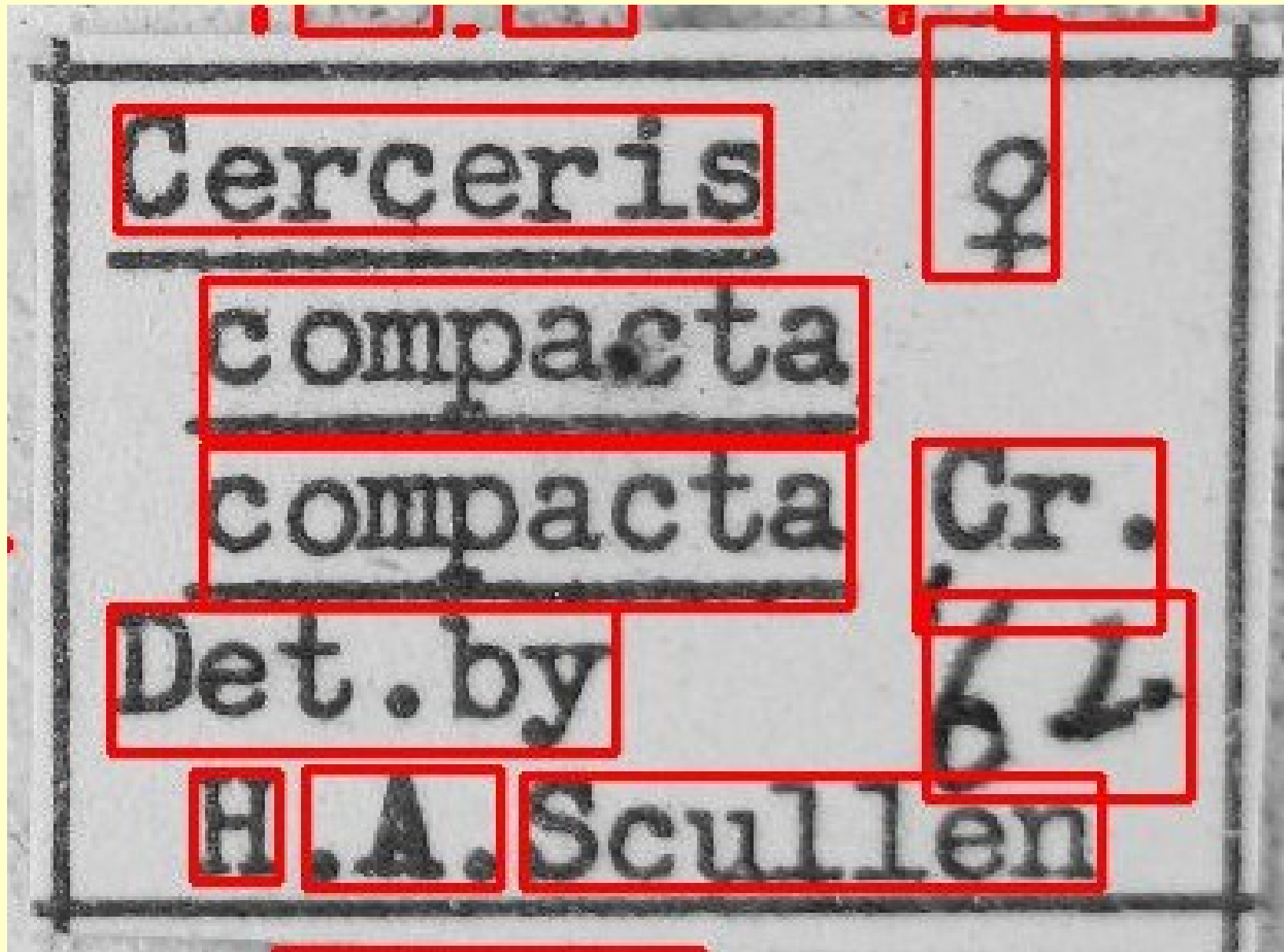
Cerceris compacta

OCR Under the Hood

HOCR output:

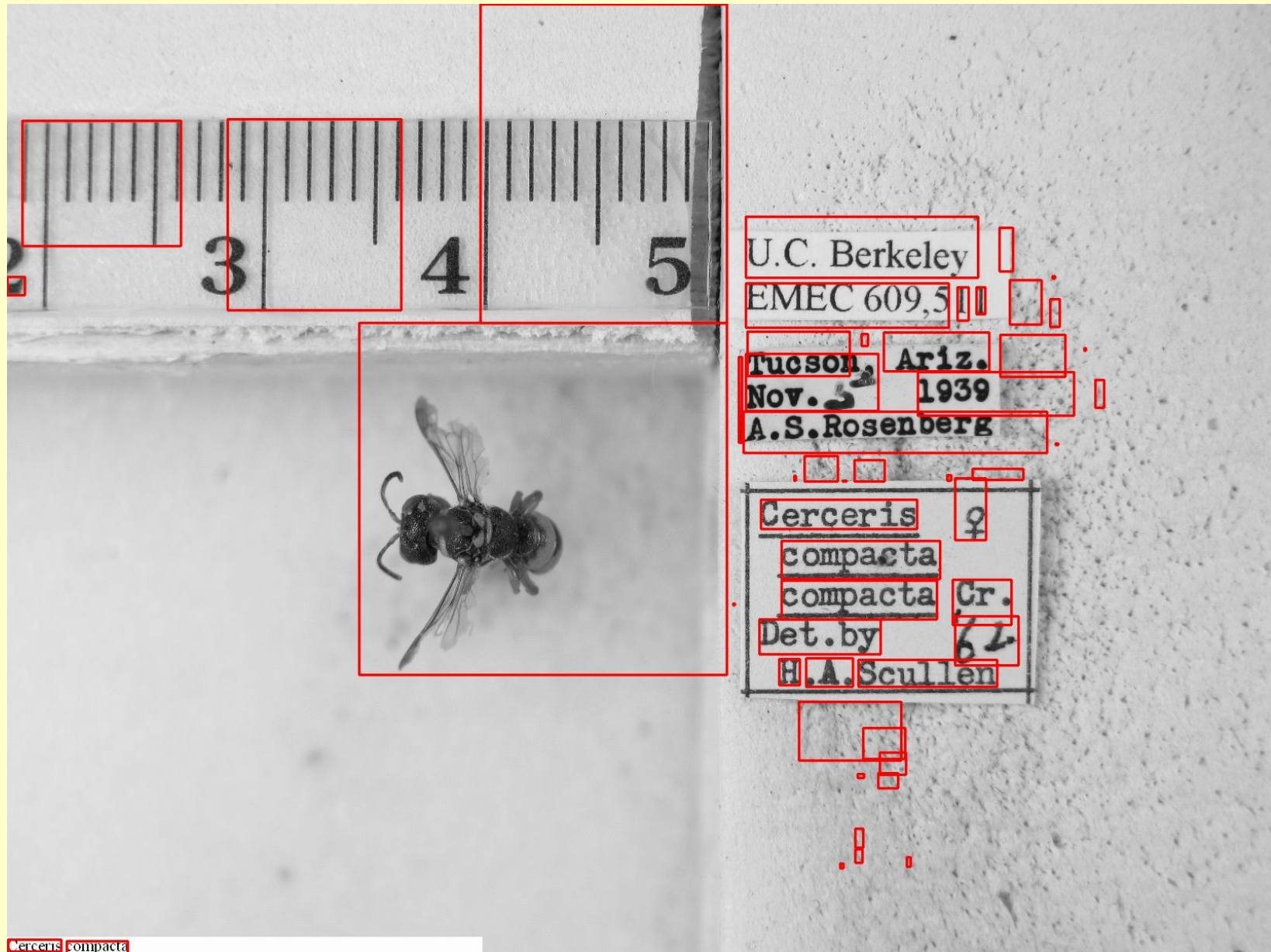
```
<span      class='ocr_word'
          id='word_7'
          title="bbox 75 1182 152 1199">
compacta
</span>
```


OCR Under the Hood



The Problem With OCR

The Problem With OCR



The Problem With OCR

Character Accuracy vs. Word Accuracy

California Digital Newspaper Collection

Character accuracy = 89%

Courtesy: Frederick Zarndt

The Problem With OCR

Character Accuracy vs. Word Accuracy

California Digital Newspaper Collection

Character accuracy = 89%

Word accuracy = character accuracy^{word length}

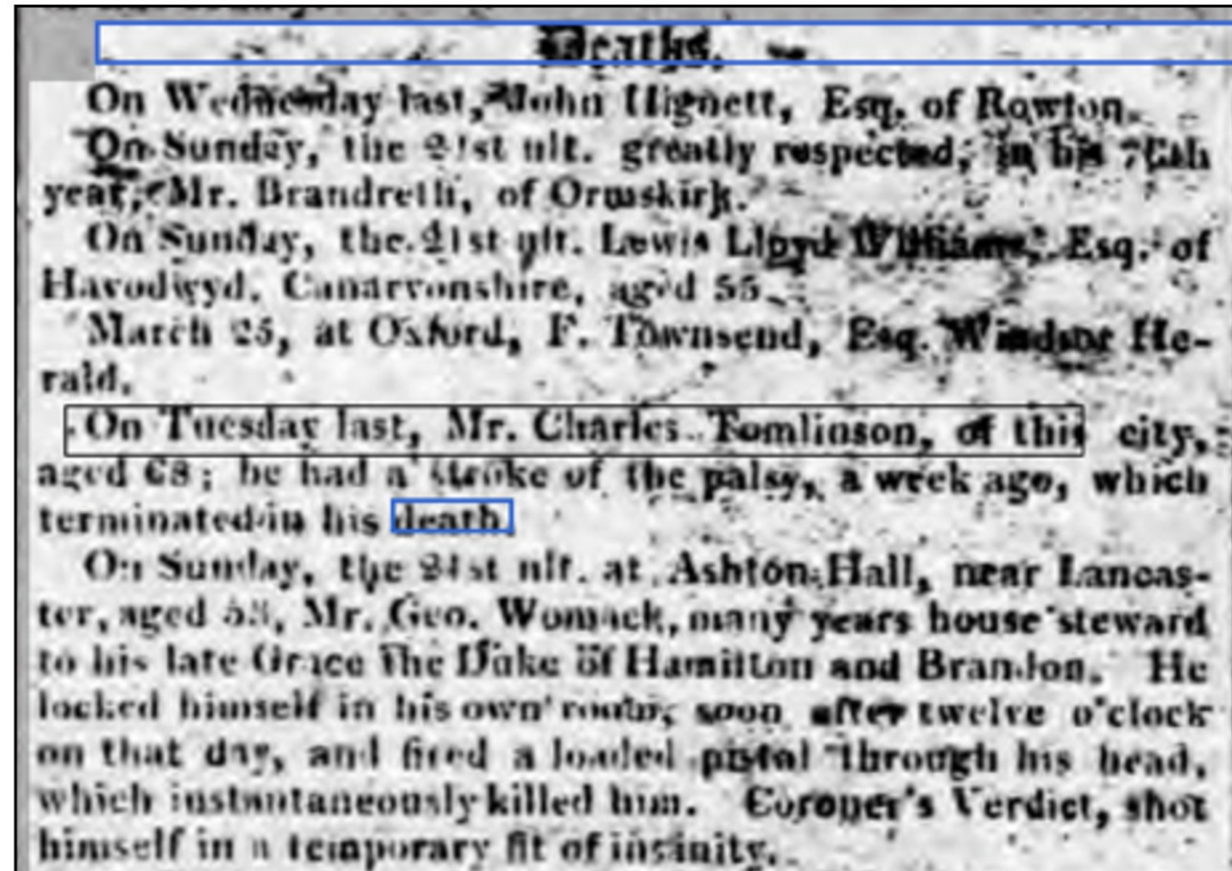
$.89^5 = 55.8\%$

Courtesy: Frederick Zarndt

raw OCR text

Deaths. **Il»rieff**, Esq. of <c .. Qn. Sunday, the till. greatly **Drandrellt**, of **Orms4\irJi.** ~ ; ;✓ ' • * On **ijfr r inn** **ljjjil F iij '11 f Havodivyd**, **Carnarvonshire**, S ; **" *- ' « ' March **Oxford, F. Tfovmeud, Uerald.** » • V .
•On Tncsdav last, Mr. **Charles. IWilinson**, this 8 ; had vf thesis#,, a week ago, which terminate<i'iu his death. . / ' ■ O'i Sunday, dJst nit. at. **AsbtCnvHall**, mar **Lancaster**, Mr., **Geo. Worn ick**, many years house'steward hit late Once The **Hamilton** and **Brandon**. He locked himself h»oWn'r«wte<: soon. twelve o'clock" that dny, and fii»-d a loaded pistol "through Ins bead, 1 which instantaneously killed him. Coronet's Verdict, shot himself in a temporary fit of Friday week,

newspaper image



Excerpt from The British Newspaper Archive, Chester Courant, Tuesday 6-Apr-1819, page 3.

Courtesy: Frederick Zarndt



uncorrected OCR accuracy by newspaper title

Title	OCR character accuracy	~OCR word accuracy
PRP Pacific Rural Press 1871 - 1922	92.6%	68.1%
SFC San Francisco Call 1890 - 1913	92.6%	68.1%
LAH Los Angeles Herald 1873 - 1910	88.7%	54.9%
LH Livermore Herald 1877 - 1899	88.6%	54.6%
DAC Daily Alta California 1841 - 1891	88.2%	53.4%
CFJ California Farmer and Journal of Useful Sciences 1855 - 1880	86.5%	48.4%
SN Sausalito News 1885 - 1922	70.4%	17.3%

*Word accuracy assumes average word length is 5 characters

Lab Session

Wikisource

OCR: Print vs. Script

W. W. CALKINS' NORTH AMERICAN LICHENS.

No. 226, *Gypsethelium*
ochroleucum, var.
hallescens, Nyl. Isl. Fla
Locality Ft. George
On Shrubs

W. W. CALKINS' NORTH
AMERICAN LICHENS.

yields

**W. W. CALKINS' NORTH
AMERICAN LICHENS '**

224, *Tryptothellium*
ochroleucum, var
hallescens, Nyl. Isl. Fla
lity, Ft. George Isl. Fla

yields

— / 2 L -
` / 17 flfl / v ; / r , 1 `
\$ / /

Implications

Methods: Humans vs. Software

Goals: Accuracy vs. Findability

Workflow: Context vs. Granularity

Findability

WEDNESDAY, OCTOBER 23, 1918

296 days past

69 to come

A good day. There was a killing
froast this morning. Ben worked
on Nat's woodings car.
Nat come tonight and he got
his car. Ben went to
Stacola this evening to see
a boat getting some Lum lumber.
a boat Mr Owens House.

Workflow

◀ BACK TO FRONT PAGE

POST SCORE TO MY WALL

Block...

arfipäivä:

artipäivä:

81

81

>>

127.

127.

Level: 1

Score: 0

Show help

= Aa Uu Ää Ää Öö f f f l l l

DIGITALKOOT

BACKGROUND

MEDIA

GIVE FEEDBACK

TERMS & PRIVACY

CONTACT

Current project

The current project consists of the National Library's archive of the issues of the newspaper [Aamulehti](#) from the end of the 19th century.

So far 93,460 people have visited the Digitalkoot web site. Volunteers have contributed a total of 266,288 minutes (5,481,490 microtasks) of their time.

DIGI TALKOOT

Digitalkoot on Facebook

Like 4,642

DIGI TALKOOT

Digitalkoot

Digitalkoot wins the Press Jury Prize for most innovative product at MindTrek LaunchPad 2011!

Lupaavia it-yrityksiä palkittiin Tampereella - Tietokone

www.tietokone.fi

Tampereella keskiviikosta perjantaille järjestetty MindTrek-tapahtuma ei houkutellut aivan yhtä paljon osallistujia kuin viime vuonna, mutta järjestäjä...

October 1 at 1:56pm

Workflow

page

transcribe

versions

[Previous Page](#)

[Next Page](#)

 [zoom in](#)

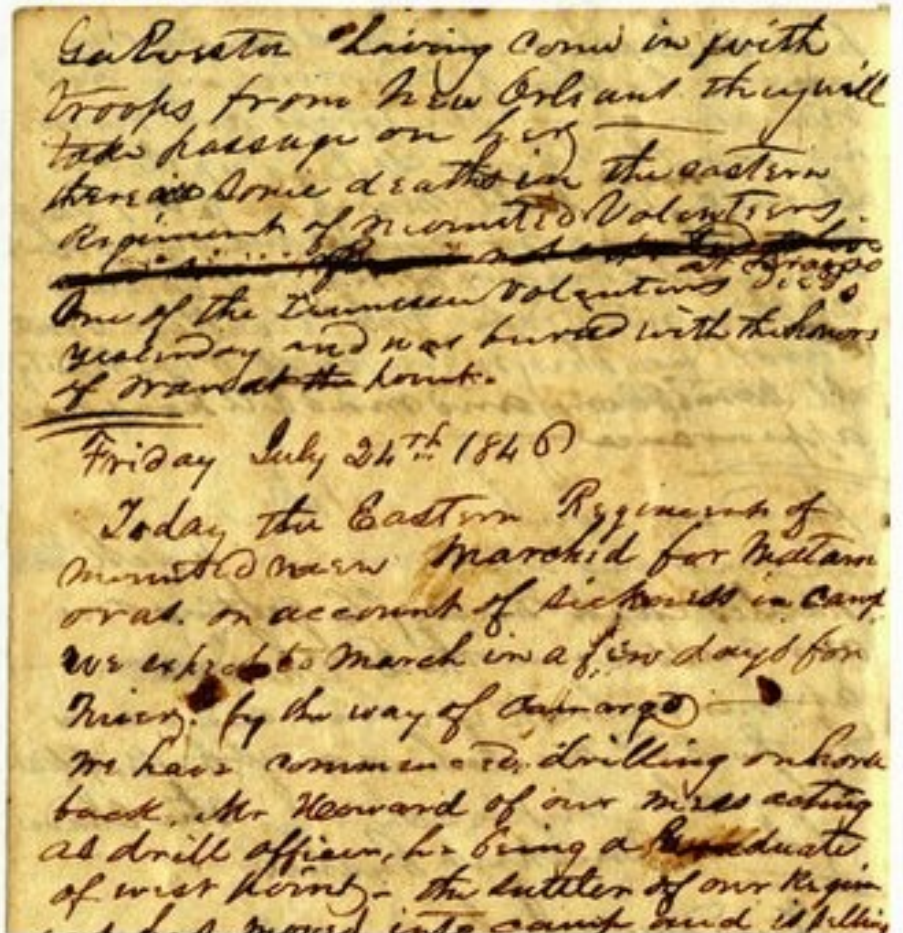
[too small?](#)

19

Galveston having come in with troops from [New Orleans](#) they will take passage on her. There are some deaths in the eastern Regiment of Mounted Volunteers One of the [Tennessee](#) Volunteers at [Brazos](#) died yesterday and was buried with honors of war at the [point](#).

Friday July 24th 1846

Today the Eastern Regiment of Mounted Men Marched for [Mata moras](#) on account of sickness in camp we expect to March in a few days for Mier by way of [Camargo](#). We have commenced drilling on horse back. [Mr Howard](#) of our mess acting as drill officer he being a Graduate



Galveston having come in with troops from New Orleans they will take passage on her. There are some deaths in the eastern Regiment of Mounted Volunteers. One of the Tennessee Volunteers at Brazos yesterday and was buried with honors of war at the point.

Friday July 24th 1846

Today the Eastern Regiment of Mounted Men Marched for Matamoras on account of sickness in camp we expect to March in a few days for Mier by the way of Camargo. We have commenced drilling on horse back. Mr Howard of our mess acting as drill officer, he being a Graduate of West Point. The sutler of our Regiment has moved into camp and is filling

Workflow

9/29/2011	Available
11/23/2011 (AM)	Transcription
11/23/2011 (PM)	Revision
12/8/2011	6 Revisions
12/9/2011	2 Revisions
8/5/2012	Revision

All by the same volunteer!

Lab Session

FromThePage

Transcribe exactly what you see!
Click image to rotate if necessary

Species/Name of Bird

Common or Scientific Name

AOU Number

Observer

Observer Name

Location

Country

State/Province

City/Town

County/Parish

Name of bird		Rosebreasted Grosbeak				
Locality		Racine, Wisconsin				
Observer		Lulu M. Lunn				
First seen	Number seen	Next seen	Became common	Last seen	Breeds	Winters
				9/13	tol.com.	no

Form BI-801

GPO

8-5280

Event Observation Data

Reference Year


	Event Month	Day	No. of Birds
First Arrival	<input type="text"/>	<input type="text"/>	<input type="text"/>
Next Seen	<input type="text"/>	<input type="text"/>	<input type="text"/>
Most Common	<input type="text"/>	<input type="text"/>	<input type="text"/>
Last Seen	<input type="text"/>	<input type="text"/>	<input type="text"/>
Breeds?	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Unknown		
Overwinters?	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Unknown		
Commonness?	<input type="text" value="Not Given"/>		

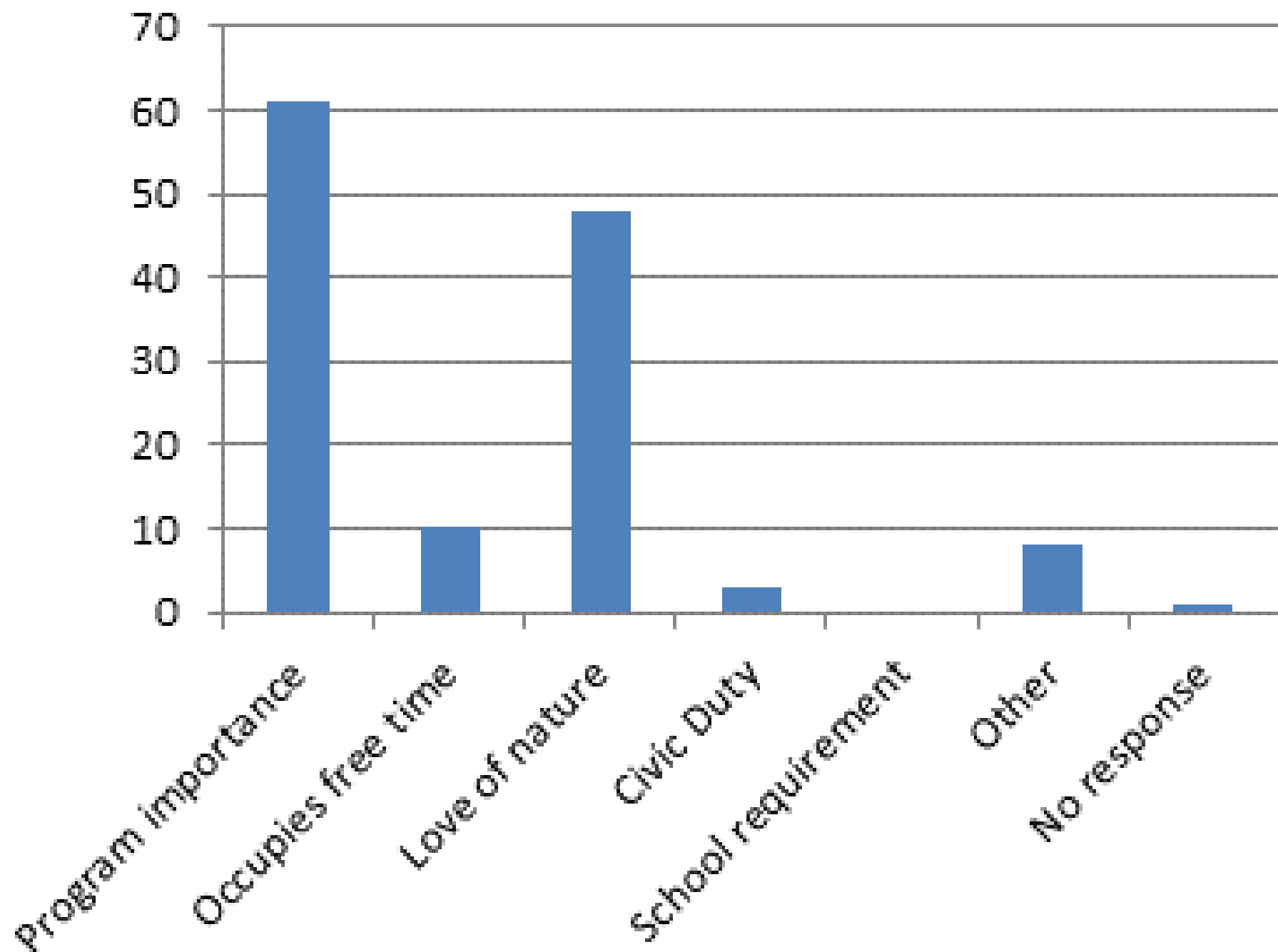
Additional Notes

Problem with Card?

Observer Notes

Transcriber Comments

 Submit



Transcribe exactly what you see!
Click image to rotate if necessary

Species/Name of Bird

Common or Scientific Name

AOU Number

Observer

Observer Name

Location

Country

State/Province

City/Town

County/Parish

Name of bird		Rosebreasted Grosbeak				
Locality		Racine, Wisconsin				
Observer		Lulu M. Lunn				
First seen	Number seen	Next seen	Became common	Last seen	Breeds	Winters
				9/13	tol.com.	no

Form BI-801

GPO

8-5280

Event Observation Data

Reference Year

	Event Month	Day	No. of Birds
First Arrival	<input type="text"/>	<input type="text"/>	<input type="text"/>
Next Seen	<input type="text"/>	<input type="text"/>	<input type="text"/>
Most Common	<input type="text"/>	<input type="text"/>	<input type="text"/>
Last Seen	<input type="text"/>	<input type="text"/>	<input type="text"/>
Breeds?	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Unknown		
Overwinters?	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Unknown		
Commonness?	<input type="text" value="Not Given"/>		

Additional Notes

Problem with Card?

Observer Notes

Transcriber Comments


 Submit



Fig. 11. The two Georges, Cantwell and Willett, at Mount Pinos, May, 1937.