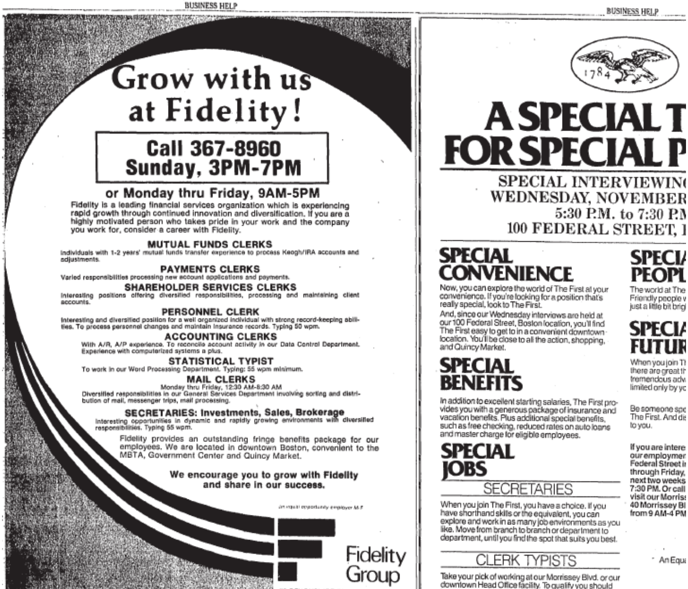
Using Text as Data

Which pages are those of job ads?

Display Ad 133 -- No Title
Boston Globe (1960-1985); Nov 4, 1979; ProQuest Historical Newspapers: The Boston Globe
pg. E51



BUSINESS HELP

Grow with us at Fidelity!

Call 367-8960
Sunday, 3PM-7PM

or Monday thru Friday, 9AM-5PM

Fidelity is a leading financial services organization which is experiencing rapid growth through continued innovation and diversification. If you are a highly motivated person who takes pride in your work and the company you work for, consider a career with Fidelity.

MUTUAL FUNDS CLERKS

Individuals with 1-2 years' mutual funds transfer experience to process Keogh/IRA accounts and adjustments.

PAYMENTS CLERKS

Varied responsibilities processing new account applications and payments.

SHAREHOLDER SERVICES CLERKS

Interesting positions offering diversified responsibilities, processing and maintaining client accounts.

PERSONNEL CLERK

Interesting and diversified position for a well organized individual with strong record-keeping abilities. To process personnel changes and maintain insurance records. Typing 50 wpm.

ACCOUNTING CLERKS

With A/R, A/P experience. To reconcile account activity in our Data Control Department. Experience with computerized systems a plus.

STATISTICAL TYPIST

To work in our Word Processing Department. Typing: 55 wpm minimum.

MAIL CLERKS

Monday thru Friday, 12:30 AM-8:30 AM

Diversified responsibilities in our General Services Department involving sorting and distribution of mail, messenger trips, mail processing.

SECRETARIES: Investments, Sales, Brokerage

Interesting opportunities in dynamic and rapidly growing environments with diversified responsibilities. Typing 55 wpm.

Fidelity provides an outstanding fringe benefits package for our employees. We are located in downtown Boston, convenient to the MBTA, Government Center and Quincy Market.

We encourage you to grow with Fidelity and share in our success.

An equal opportunity employer M/F

Fidelity Group

BUSINESS HELP

A SPECIAL T
FOR SPECIAL P

SPECIAL INTERVIEWIN
WEDNESDAY, NOVEMBER
5:30 P.M. to 7:30 P.M
100 FEDERAL STREET, I

SPECIAL CONVENIENCE

Now, you can explore the world of The First at your convenience. If you're looking for a position that's really special, look to The First.

And, since our Wednesday interviews are held at our 100 Federal Street, Boston location, you'll find The First easy to get to in a convenient downtown location. You'll be close to all the action, shopping, and Quincy Market.

SPECIAL BENEFITS

In addition to excellent starting salaries, The First provides you with a generous package of insurance and vacation benefits. Plus additional special benefits, such as free checking, reduced rates on auto loans and master charge for eligible employees.

SPECIAL JOBS

SECRETARIES

When you join The First, you have a choice. If you have shorthand skills or the equivalent, you can explore and work in as many job environments as you like. Move from branch to branch or department to department, until you find the spot that suits you best.

CLERK TYPISTS

Take your pick of working at our Morrissey Blvd. or our downtown Head Office facility. To qualify you should

SPECIA PEOPL

The world at The
Friendly people w
just a little bit brig

SPECIA FUTUR

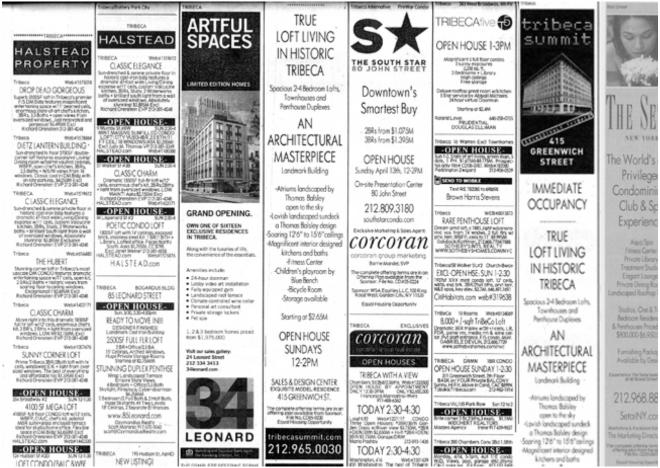
When you join Th
there are great th
tremendous adv
limited only by yo

Be someone spe
The First. And dis
to you.

**If you are intere
our employmen
Federal Street i
through Friday,
next two weeks
7:30 PM. Or call
visit our Morris
40 Morrissey Bl
from 9 AM-4 PM**

An Equa

Which pages are those of job ads?



TRIBECA

HALSTEAD PROPERTY

Tribeca Web#1375018

DROP DEAD GORGEOUS

Superb 3000SF loft in Tribeca's premier F/S DM Bldg features magnificent entertaining space w/11' beamed ceils, enormous state-of-art chef's kitchen, 3BRs, 3.5Baths + open views from oversized windows. Just renovated and gorgeous! [illegible] Excl
Richard Orenstein 212-381-4248

Tribeca Web#1378864

DIETZ LANTERN BUILDING

Sun-drenched hi floor 3700SF double-corner loft features expansive Living/Dining room w/barrel vaulted ceilings, WBFP, open chef's kitchen, 3BRs, 2.5 Baths + N/S/W views from 18 windows. Classic cool in DM Bldg with on-site parking. $4.25M Excl
Richard Orenstein EVP 212-381-4248

TRIBECA Web#1579612

CLASSIC ELEGANCE

Sun-drenched & serene private floor in historic cast-iron bldg features a dramatic 47-foot wide Living/Dining expanse w/11' ceils, custom Valcucine kitchen, 3BRs, Study, 2 Waterworks baths + brilliant South light from a wall of oversized windows. Absolutely stunning! $2.85M Exclusive
Richard Orenstein EVP 212-381-4248

Tribeca Web#[illegible]

THE HUBERT

Stunning corner loft in Tribeca's most upscale DM CONDO features dramatic entertaining space w/11' ceils, open kit, 2 BRs/2 Baths + historic views from soaring floor to ceiling windows. Exceptional! [illegible] Excl
Richard Orenstein EVP 212-381-4248

Tribeca Web#1403171

CLASSIC CHARM

Move right into this dramatic 1850SF full flr loft w/12' ceils, enormous chef's kit, 2 BR's, 2 Bths + light from oversized windows. LOW MT $2.150M Excl
Richard Orenstein EVP 212-381-4248

Tribeca Web#1007676

SUNNY CORNER LOFT

Prime Tribeca 2BR/2Bath loft with hi ceils, windowed EIK + light from oversized windows. The best of everything and affordable too $1.[illegible] Excl
Richard Orenstein EVP 212-381-4248

-OPEN HOUSE-

[illegible] #2 SUN 12-1:30

4100 SF MEGA LOFT

4100SF full floor CONDO loft w/12' ceils, WBFP, C/A/C, chef's kit, palatial MBR suite+glass enclosed terrace ideal for studio/home office. Flexible space in DM Bldg! $3.[illegible]M Excl
Richard Orenstein EVP 212-381-4248
HALSTEAD.com Web#1442330

-OPEN HOUSE-

[illegible] Hudson St #2D SUN 12-1:30

LOFT CONDO/BALC [illegible]

Tribeca/Battery Park City

TRIBECA

HALSTEAD

TRIBECA Web#1519613

CLASSIC ELEGANCE

Sun-drenched & serene private floor in historic cast-iron bldg features a dramatic 47-foot wide Living/Dining expanse w/11' ceils, custom Valcucine kitchen, 3BRs, Study, 2 Waterworks baths + brilliant south light from a wall of oversized windows. Absolutely stunning! $2.85M Excl
Richard Orenstein EVP 212-381-4248

-OPEN HOUSE-

[illegible] SUN 2:30-4

MINT MASSIVE SUNFILLED CONDO LOFT! CITY VIEWS! [illegible] FT CEIL! 18 WINDOWS! ASK $2.250M! Excl Judy M. Thomas VP 212-381-3349
HALSTEAD.com Web#1100388

-OPEN HOUSE-

57 Walker St #3B SUN 2:30-4

CLASSIC CHARM

Dramatic 1850SF full-flr loft w/12' ceils, enormous chef's kit, 2BRs/2Bths + light from oversized windows. LOW MAINT! Asks $2.150M Excl
Richard Orenstein EVP 212-381-4248
HALSTEAD.com Web#1403171

-OPEN HOUSE-

[illegible] Lispenard St #2 SUN 2:30-4

POETIC CONDO LOFT

1800SF loft with 14' ceilings, exposed brick, stainless steel Kit, 1 BR/1 BTH + Library, Lofted office. Faces North/South. Asks $1.[illegible]M, CC $[illegible]
Excl Janet Weiner 212-381-[illegible]
HALSTEAD.com Web#1575876

HALSTEAD.com

TRIBECA BOGARDUS BLDG

85 LEONARD STREET

-OPEN HOUSE-

Sun. [illegible], 2:30-4:30pm

READY TO MOVE IN!!!

DESIGNER FINISHES!
Landmark Cast Iron Building

2500SF FULL FLR LOFT

2 BR+Office/2.5 BA
15' Ceilings, Arched Windows, Huge Private Storage Rooms
Starting at $2.75MM

STUNNING DUPLEX PENTHSE

Wrap Landscaped Terrace
Empire State Views,
4 Bedroom + Office/3.5 Bath
Skylight, Fireplace, Cyber-doorman
$[illegible]
3 Bedroom/2 Full Bath & 2 Half Bath,
Huge Skylights At Two Levels
18' Ceilings, 2 Separate Entrances

www.85Leonard.com

Ozymandias Realty
Scott Morano 917-570-[illegible]
scott@OzymandiasRealty.com

TRIBECA 195 Hudson St, Apt 4D

NEW LISTING!

TRIBECA

ARTFUL SPACES

LIMITED EDITION HOMES

GRAND OPENING.

OWN ONE OF SIXTEEN EXCLUSIVE RESIDENCES IN TRIBECA.

Along with the luxuries of life, the convenience of the essentials.

Amenities include:

- 24-hour doorman
- Lobby video art installation
- Fully equipped gym
- Landscaped roof terrace
- Climate-controlled wine cellar
- Personal art consultant
- Private storage lockers
- Pet spa

1, 2 & 3 bedroom homes priced from $1,975,000

Visit our sales gallery:
24 Leonard Street
212 334 3413
34leonard.com

34 LEONARD

Marketing and Exclusive Sales Agent: The Marketing Direction, Inc.

THE COMPLETE OFFERING TERMS [illegible]

TRUE LOFT LIVING IN HISTORIC TRIBECA

Spacious 2-4 Bedroom Lofts, Townhouses and Penthouse Duplexes

AN ARCHITECTURAL MASTERPIECE

Landmark Building

-Atriums landscaped by Thomas Balsley open to the sky
-Lavish landscaped sundeck a Thomas Balsley design
-Soaring 12'6" to 15'6" ceilings
-Magnificent interior designed kitchens and baths
-Fitness Center
-Children's playroom by Blue Bench
-Bicycle Room
-Storage available

Starting at $2.65M

OPEN HOUSE SUNDAYS 12-2PM

SALES & DESIGN CENTER
EXQUISITE MODEL RESIDENCE
415 GREENWICH ST.

The complete offering terms are in an offering plan available from Sponsor. File No. CD05-0030
Equal Housing Opportunity

tribecasummit.com
212.965.0030

Tribeca Alternative PreWar Condo

THE SOUTH STAR
80 JOHN STREET

Downtown's Smartest Buy

2BRs from $1.075M
3BRs from $1.395M

OPEN HOUSE
Sunday April 13th, 12-2PM

On-site Presentation Center
80 John Street

212.809.3180
southstarcondo.com

Exclusive Marketing & Sales Agent:

corcoran

corcoran group marketing
Barrie Mandel, SVP

The complete offering terms are in an Offering Plan available from the Sponsor. File No. CD#05-0224

Sponsor: WSA Equities LLC, 100 Ring Road West, Garden City, NY 11530

Equal Housing Opportunity

TRIBECA EXCLUSIVES

corcoran
corcoran group real estate

OPEN HOUSES

TRIBECA WITH A VIEW

Chambers St/2bd/2.5bths Web#1233000
OPEN HOUSE BY APPOINTMENT ONLY 12:30-2PM ONLY $2,650,000
Francesca Mannarino-Wertz
917-488-6362

TODAY 2:30-4:30

Laight/[illegible] Web#1207117 CONDO
Three Open Houses: *2BR/2Bth Garden Oasis w/River view $1.750M, *3BR +DEN/3Bth $2.[illegible]M + [illegible] *2BR+DEN/[illegible] $2.750M, Garage/DRM
Iliana Pashby 212-893-1436

TODAY 2:30-4:30

Washington, 416 Web#1201[illegible]
[illegible] Washington. The heart of Tribeca [illegible]

Tribeca 283 West Broadway, 6th Flr

TRIBECAfive

OPEN HOUSE 1-3PM

Magnificent 5 full floor condos
3 sunny exposures
2,250 sq. ft.
3 Bedrooms + Library
High ceilings
Free storage

Deluxe rooftop great room w/kitchen
5 Star service by Abigail Michaels
24 hour virtual Doorman

Starting at $2.4M

Roland Levin 646-258-0755
PRUDENTIAL DOUGLAS ELLIMAN

Tribeca 16 Warren Excl Townhomes

-OPEN HOUSE-

Sun 1-3: [illegible]
Paddington Zwigard 212-906-0539

SEND TO MOBILE

Text RE 78[illegible] to [illegible]

Brown Harris Stevens

Tribeca WEB#[illegible]

RARE PENTHOUSE LOFT

Dream grnd loft, [illegible] light w/panoramic vus from 14 wndws, 2 full flrs w/ priv. terr, WBFP, ceils to 11' $[illegible]
Goldbeck/Kottman, 212.606.7704/7688
SOTHEBY'S INT'L REALTY
WWW.SOTHEBYSHOMES.COM/NYC

Tribeca/58 Walker St #2 Church-Bway

EXCL-OPEN HSE: SUN 1-2:30

[illegible] XXX mint condo loft, 12' ceils, [illegible]
CitiHabitats.com web#319638

TriBeCa 10 Rooms Web#[illegible]

8,000+/-sqft TriBeCa Loft

Dramatic 3BR triplex w/20+ft ceils, LR, FDR, game rm, media rm & wine cellar. Pvt staff entrance. F/S condo. [illegible]
GABRIELE DEVLIN, 212.606.7729
www.sothebyshomes.com/nyc

TRIBECA DRMN 1BR CONDO

OPEN HOUSE SUNDAY 1-2:30

201 Greenwich Street, 7th Floor
BASK on YOUR Private BALCONY
Sunny, Hi Flr, Move in Cond, CAC $[illegible]
Tribeka Tribeca.com 212-962-1914

Tribeca Vic,165 Park Row Sun 12 to 2

-OPEN HOUSE-

[illegible] $1.1M
WEICHERT REALTORS
Mazzeo Agency Irene 917-[illegible]

Tribeca 200 Chambers Conv 2Bd 1.5Bth

-OPEN HOUSE-

Monday, 4/14, 5-7pm, Apt 11E condo
W/D, Views, [illegible]

Tribeca

tribeca summit

415 GREENWICH STREET

IMMEDIATE OCCUPANCY

TRUE LOFT LIVING IN HISTORIC TRIBECA

Spacious 2-4 Bedroom Lofts, Townhouses and Penthouse Duplexes

AN ARCHITECTURAL MASTERPIECE

Landmark Building

-Atriums landscaped by Thomas Balsley open to the sky
-Lavish landscaped sundeck a Thomas Balsley design
-Soaring 12'6" to 15'6" ceilings
-Magnificent interior designed kitchens and baths

Wall Street

THE SE

NEW YORK

The World's
Privilege
Condomini
Club & Sp
Experienc

Aqua Spa
Fitness Center
Private Library
Treatment Studi
Elegant Lounge
Private Dining Ro
Landscaped Rooftop

Studios, One & T
Bedroom Residen
& Penthouses Priced
$800,000-$6,950,0

Furnishing Packag
Available by Desi

Experience The Setai Sale
at 40 Broad Street

212.968.88

SetaiNY.co

Marketing & Exclusive Sa
The Marketing Directi

The complete offering term
offering plan from the sp
File No. CD05-[illegible]

Equal Housing Opportu

Topic classification

- There are a number of possible types of pages of advertisements: for jobs, for real estate, for cars, in retail, etc...
- Within each type of advertisements certain types of words are more likely to appear:
  - employment, responsibility, manager in job ads
  - foot, loft, architectural in real estate ads
- If we knew the predictive words for each type of ad, we could estimate the likelihood for each topic for each page.
- Goal: Figure out the "predictive words" of each type of ad.
- Idea: Words come from different *topics.* Within a given page of ads, certain topics are more or less represented.

Latent Dirichlet Allocation

Let

- $V =$ size of vocabulary; $K =$number of topics
- α $(K \times 1)$ and β $(K \times V)$ parameters to be estimated
- w (a word) is a vector of length V. $w_i = 1$ for one i and 0 elsewhere.
- d_m is a particular page of ads.

Latent Dirichlet Allocation

Each page of ads comes from a mixture of topics:

- For page m, random variables θ_{mk} are drawn from a (Dirichlet-α) distribution.
- For each word within a page, the probability that a word is from topic $z = z_k$ is θ_{mk}
- Conditional on the topic, $\beta_{ki} = \Pr\left(w_i = 1 | z = z_k\right)$
- We can write the likelihood of each doc:

$$\begin{aligned}\Pr\left(d_m | \alpha, \beta\right) &= \int p\left(\theta | \alpha\right) \cdot \times \\ &\left(\prod_{n \in \text{Words in } d_m} \sum_{k \in \text{Topics}} \Pr\left(z_k | \theta\right) \Pr\left(w_n | z_k, \beta\right)\right) d\theta\end{aligned}$$

Latent Dirichlet Allocation

- Need to reduce the dimensionality of the maximization problem.
  - Stem
  - Drop very frequent and infrequent words.

Results for $K = 3$ (highest β word stems for each topic):

- Topic 1: new, home, owner, acr, call, car, hous, den, area, ask
- Topic 2: resum, seek, call must, work, exp, excel, new, salari, send.
- Topic 3: build, ave, new, park, call, studio, east, avail, fee, firm

Similarity

- Do the words/phrases lpn, rn, registered nurse, nurse represent the same occupation?
  - What about finance professor and econometrician?
  - What about finance professor and nurse?
- Do the words/phrases delegating and directing refer to the same work activity?
  - What about monitor and motivating?
  - What about scheduling and coordinating?

Idea: Similar words appear in similar contexts:

- "Limits how many patients can be assigned to each *registered nurse* in Massachusetts hospitals and certain other healthcare facilities."
- "Newton-Wellesley Hospital *registered nurse* Betty Sparks is a member of the Committee to Ensure Safe Patient Care, which advocates for increased staffing levels."
- "With a busy hospital job as a *licensed practical nurse*, it was hard not to miss some of those frequent injections."

Continuous Bag of Words

- Suppose we see a chunk of text that is of the form " each registered X Massachusetts hospitals" What is the probability that...
    - X="nurse"
    - X="econometrician"
    - X="programmer" ?
- Would like representation of words that assigns a high probability to X=nurse.
- Model to estimate $\mathcal{C}$, $\tilde{\mathcal{C}}$ each which is of dimension V (number of words) $\times$ n (number of "features")
- As before: w (a word) is a vector of length V. $w_i = 1$ for one i and 0 elsewhere.

Continuous Bag of Words

1. Take word vectors for each of the words in the context of (e.g., w_{nurse}, $w_{\text{registered}}$, $w_{\text{Massachusetts}}$, $w_{\text{hospitals}}$)
2. Recover the embedded word vectors for a given context: $\tilde{\mathcal{C}}' w_{\text{nurse}}$, $\tilde{\mathcal{C}}' w_{\text{registered}}$, $\tilde{\mathcal{C}}' w_{\text{Massachusetts}}$, $\tilde{\mathcal{C}}' w_{\text{hospitals}}$.
3. Average these to get a measure of the context: $\tilde{v} = \frac{\tilde{\mathcal{C}}' w_{\text{nurse}} + \tilde{\mathcal{C}}' w_{\text{registered}} + \tilde{\mathcal{C}}' w_{\text{Massachusetts}} + \mathcal{C}' w_{\text{hospitals}}.}{4}$.
4. Generate a probability of this context: $p_c = \frac{C\tilde{v}}{\sum C\tilde{v}}$

Log likelihood of a data set is given by comparing the actual words in the "X" position to what would be predicted out of step 4

$$\log \mathcal{L} = \sum_c y_c \cdot \log p_c$$

Examples Elsewhere

- Gentzkow and Shapiro (2010): Measuring media slant.
  - Link Congressional record (words/phrases used by Dem./Repub. congresspeople) to newspapers.
    - Democrats mention "estate tax," while Republicans mention "death tax"
    - Washington Post: 13.7 ratio mentions of estate tax to death tax.
    - Washington Times: 1.3 ratio mentions of estate tax to death tax.
- Baker, Bloom, Davis (2016): Measuring policy uncertainty.
- In the optional papers.
  - Hoberg and Philips: SEC Filings to learn about firms' products/competitors
  - Mann and Puttmann: Google text of all US patents to learn about new technologies which lead to automation.
- For other examples, see Gentzkow, Kelly, Taddy (2017, JEL)