The Effects of Immigration on Native Employment — Evidence from Isolated and Integrated Ethnic Enclaves in Germany

Ilya Kukaev[1], Ahmed Rahman[2]

[1]HSE University
[2]Lehigh University

June 25, 2025

Introduction

- We are interested in the different effects on native employment, wages and sentiment from immigrants with different levels of assimilation with the culture of their chosen land.
- We are interested in different measures of “assimilation.”
  - Economic
  - Linguistic
  - Cultural
  - Social

Why Does It Matter?

- Important for understanding the *causes* of migration.
  - Bartik instruments may capture a sub-set of migrant types.
- Important for understanding the *consequences* of migration.
  - Migrants with different levels of assimilation may differently affect native wage distributions (separate from education levels).

Introduction

- We will use Google Trends data to assess the 1)linguistic and 2) cultural assimilation of immigrants populations.
  - This paper — assesses *linguistic* assimilation for migrant groups across NUTS2 regions in Germany.
  - Future work — will assess more broader cultural assimilation for migrant groups across NUTS2 in Europe and MSAs in the United States.

Some Potential Findings

- Less culturally assimilated immigrants may push natives out of less educated or more manual jobs [Peri (2016), Peri and Sparber (2009), Dustmann, Fabbri & Preston (2005)].
- Less culturally assimilated immigrants may generate more cultural goods and services to the local economy (Portes Wilson (1980)).
- Less culturally assimilated immigrants may produce economic enclaves (Borjas 2000).

Motivation

- We redefine the shares in a shift-share instrument using Google Trends and several correction factors.
  - Now instrument is designed to distinguish relatively isolated enclaves from more integrated enclaves.
- We use The Data for Integration Project (D4I) dataset.

Findings in this Paper

- A 10% increase in the relative supply of immigrants in linguistically assimilated areas increases growth in log native employment on average by 0.16 percentage points.

- A 10% increase in the relative supply of immigrants in linguistically unassimilated areas does not produce any contemporaneous growth in log native employment.

Rest of the Talk

- Some literature
- Conceptual model
- Numerical example of the Dis-assimilation Index
- Methodology
- Data and Results
- Future work

Some Literature

Some Literature

1. Ethnic enclaves
   - Size of the enclave ↓ host country language acquisition and earnings (Chiswick and Miller 2005)
   - The size has ↑ on earnings for refugees (Edin et al. (2001), Damm (2009))
   - Higher probability of employment and lower human capital investment with ↑ size (Battisti et al. 2018)
2. Language
   - Language classes have positive effects on earnings (Arendt et al. 2020)
   - Language proficiency of the host country language is correlated with earnings (Dustmann and Van Soest (2002)[1])
   - Multiple equilibria in language learning (Brock et. al. 2022)
3. Shift-share instruments
   - Review (Goldsmith-Pinkham et al. 2020)
   - Need for lagged inflow variable in the LR (Jaeger et al. 2018)

[1] Noted measurement error in self-reporting language proficiency data

Conceptual Model

A Simple Search Model

1. Use DMP model as a starting point[2]
2. Two inter-connected labor markets — primary market where natives and migrants work, and enclave market where only migrants work.
3. In equlibrium:

$$N_1 = \underbrace{\frac{\theta_1^{1/2}}{s_1 + \theta_1^{1/2}}}_{positive \;\; effect} - \underbrace{\hat{\delta}}_{negative \;\; effect}$$

N = Native employment in the primary economy
θ = labor market tightness
s = job separation rate
$\hat{\delta}$ = share of immigrants joining primary economy

[2] Similar to Ortega (2000)

Predictions from model

- Predicts positive and negative effects of immigration on native employment
- Positive effect - immigrants have higher search costs and lower wage which allows firms to open more vacancies
- Negative effect - replacement effect

Numerical example of the Dis-assimilation Index

Simple example

- Total population of NUTS2 region is as follows

NUTS2 Assimilated?	Total n	GER	POL Yes	POL No	TUR Yes	TUR No
j	200	100	35	25	30	10
k	100	70	5	15	2	8

Simple example

- Total population of NUTS2 region is as follows

NUTS2 Assimilated?	Total n	GER	POL Yes	POL No	TUR Yes	TUR No
j	200	100	35	25	30	10
k	100	70	5	15	2	8

Simple example

- Total population of NUTS2 region is as follows

NUTS2 Assimilated?	Total n	GER	POL Yes	POL No	TUR Yes	TUR No
j	200	100	35	25	30	10
k	100	70	5	15	2	8

Simple example

- Total population of NUTS2 region is as follows

NUTS2 Assimilated?	Total n	GER	POL Yes	POL No	TUR Yes	TUR No
j	200	100	35	25	30	10
k	100	70	5	15	2	8

Simple example

- Total population of NUTS2 region is as follows

NUTS2 Assimilated?	Total n	GER	POL Yes	POL No	TUR Yes	TUR No
j	200	100	35	25	30	10
k	100	70	5	15	2	8

Simple example

- Total population of NUTS2 region is as follows

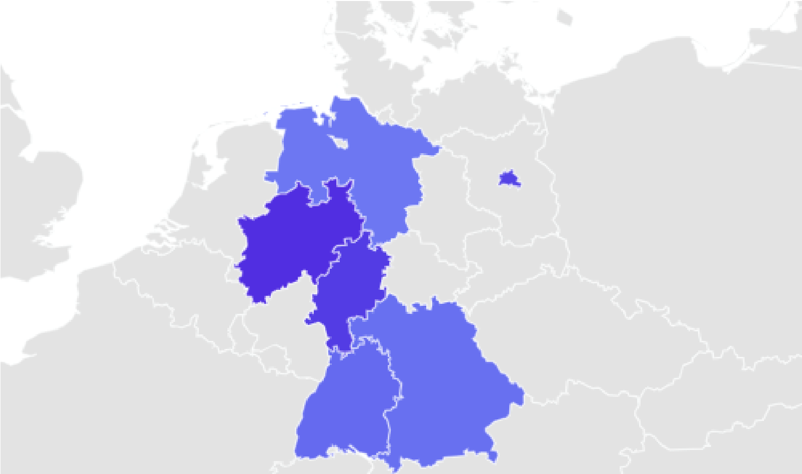
NUTS2 Assimilated?	Total n	GER	POL Yes	POL No	TUR Yes	TUR No
j	200	100	35	25	30	10
k	100	70	5	15	2	8

Dissimilarity index for a single query in immigrants' mother language

$$\alpha_{\lambda j} = h\left(\frac{25/60}{15/20}\right)$$

Details

An Example — "Wasser" vs. "Woder"



Strengths and Weaknesses

- Based on actual behavior rather than self assessment
- Potentially captures illegal immigrants too
- Some limitations:
  - Tourists are included (may incorporate seasonal corrections)
  - Cannot capture those without internet access (may incorporate controls).

Methodology

The Steps

1. Define what is behind a google rank, derive correction factors, and obtain disassimilation index (DI)
2. Define a measure for clusterisation of immigrants
3. Regress DI on the clusterisation measure
4. Obtain residuals to construct shares for shift-share instrument

Methodology 1 - The Dissimmilarity Index

$$DI_{\lambda j} = \left[\Pi_{w=1}^{Z} \alpha_{Z\lambda j}\right]^{1/Z}$$

Methodology 2 — Clusterization

Kullback Leibler divergence to capture clusterisation of immigrants

$$KL(P\|Q) = \sum_{x \in \chi} P(x) log(\frac{P(x)}{Q(x)}) \quad (1)$$

where

$Q(x)$ is the uniform distribution as if all immigrants were distributed evenly across all NUTS 2 regions

$P(x)$ is the actual distribution of immigrants in NUTS 2 regions

Methodology 3 — Regress DI on clusterization

$$DI_{\lambda jt^0} = \alpha_0 + \alpha_1 KL_{\lambda jt^0} + MIG_\lambda + e_{\lambda jt^0}$$

where

$DI_{\lambda jt^0}$ is a disassimilation index for language group λ in NUTS2 region j in 2011

MIG_λ the language-group-specific dummy variables

$KL_{\lambda jt^0}$ is the Kullback Leibler divergence that shows the clusterization of immigrants

Methodology 4 - Residuals

- With $e_{\lambda j t^0} > 0$ we presume inflows of migrants are "unassimilated."
- With $e_{\lambda j t^0} < 0$ we presume inflows of migrants are "assimilated."

Methodology 4 - Unassimilated shares

$$\frac{\hat{e}^{+}_{\lambda j t^0}}{\hat{e}^{+}_{\lambda t^0}} = \begin{cases} \frac{\hat{e}_{\lambda j t^0}}{\hat{e}_{\lambda t^0}} \; if \; \hat{e}_{\lambda j t^0} > 0 \\ 0 \; if \; \hat{e}_{\lambda j t^0} < 0 \end{cases} \tag{2}$$

where $\hat{e}^{+}_{\lambda t^0}$ are positive residuals summed over all NUTS2 j regions. Then we can redefine the standard Bartik instrument in the following form to capture unassimilated immigrant inflows:

$$\hat{m}_{jt} = \sum_{\lambda} \frac{\hat{e}^{+}_{\lambda j t^0}}{\hat{e}^{+}_{\lambda t^0}} \frac{\Delta M_{\lambda t}}{L_{jt-1}} \tag{3}$$

Methodology 4

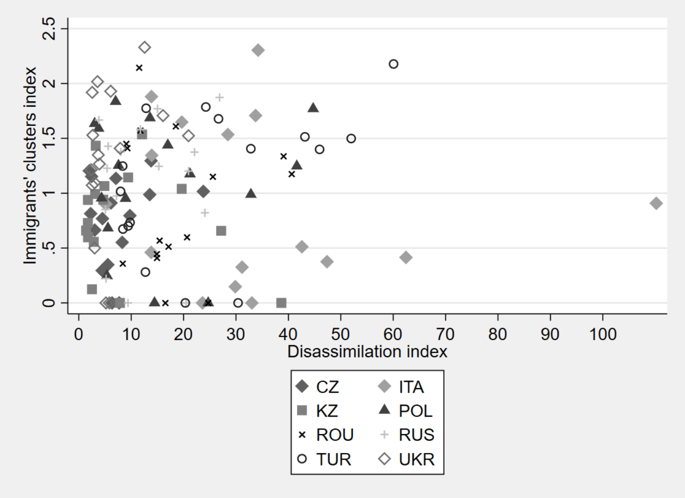
$$\frac{\hat{e}^{-}_{\lambda j t^0}}{\hat{e}^{-}_{\lambda t^0}} = \begin{cases} \frac{\hat{e}_{\lambda j t^0}}{\hat{e}_{\lambda t^0}} \;\; if \;\; \hat{e}_{\lambda j t^0} < 0 \\ 0 \;\; if \;\; \hat{e}_{\lambda j t^0} > 0 \end{cases} \tag{4}$$

where $\hat{e}^{-}_{\lambda t^0}$ are negative residuals summed over all NUTS2 j regions.

Then we can redefine the standard Bartik instrument in the following form to capture assimilated immigrant inflows:

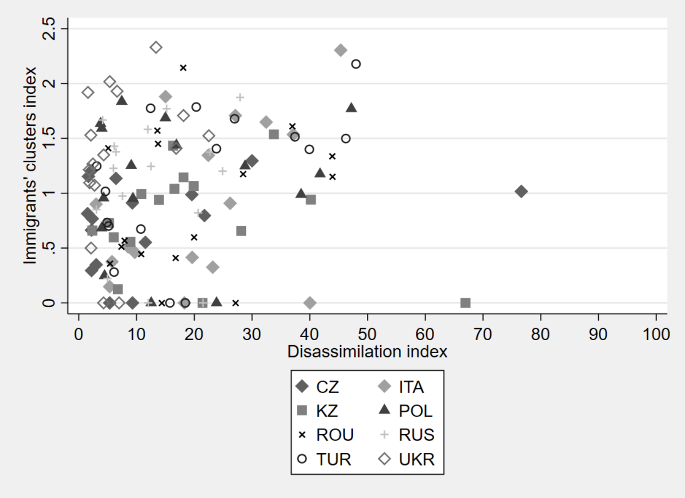
$$\hat{\hat{m}}_{jt} = \sum_{\lambda} \frac{\hat{e}^{-}_{\lambda j t^0}}{\hat{e}^{-}_{\lambda t^0}} \frac{\Delta M_{\lambda t}}{L_{jt-1}} \tag{5}$$

Methodology.



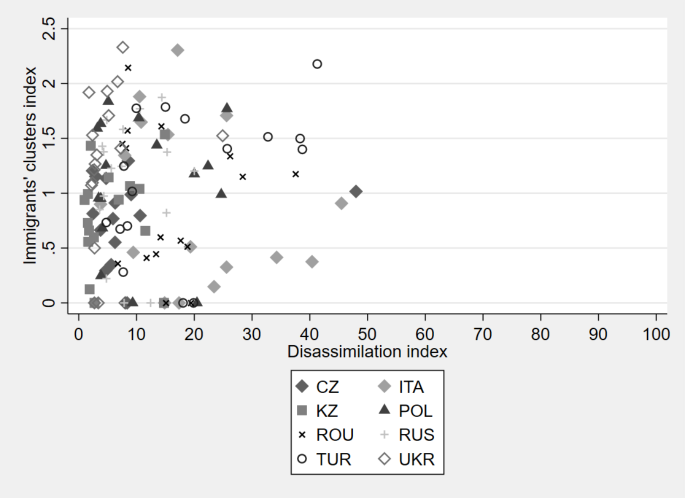
Immigrants' clustering vs isolation index for 2011. Generic queries

Methodology.



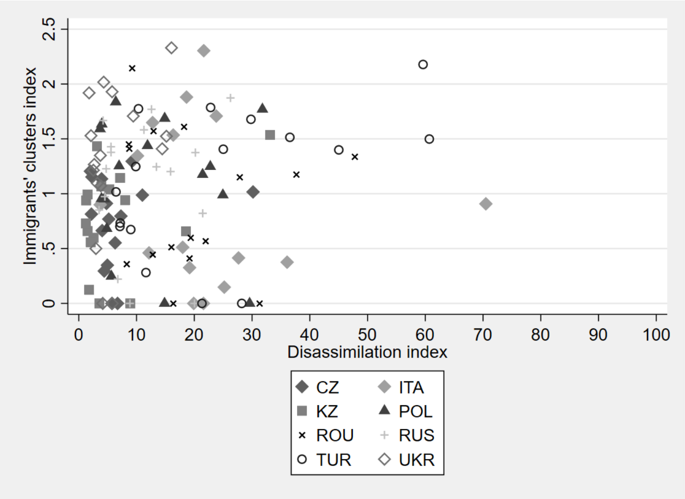
Immigrants' clustering vs isolation index for 2011. Common queries

Methodology.



Immigrants' clustering vs isolation index for Jan-Feb 2011. Generic queries

Methodology.



Immigrants' clustering vs isolation index for Nov-Dec 2011.
Generic queries

Data and Results

Geographic Breakdown

- NUTS 1 states e.g. Baden-Württemberg (DE1)
- NUTS 2 government regions e.g. Stuttgart (DE11)
- NUTS 3 districts e.g. Böblingen (DE112)
- LAU 1 collective municipalities
- LAU 2 municipalities

Data Sources

Variable	Geo level	Source
Stocks of foreigners	NUTS 2	Eurostat
Stocks of locals	NUTS 2	Eurostat
Native employment in thousand	NUTS 2	Eurostat
Inflows of immigrants by language groups	Country	OECD
Shares of immigrants by language groups	LAU 2	D4I project
Disassimilation index	NUTS 2	Google Trends

Data sources

Language Groups

Country	immigrant language groups account for, %
Polish	29.8
Turkish	14
Russian	11.6
Kazakh	8.1
Czech	4
Italian	3
Ukrainian	1.7

Proportion of all immigrants excluding nonidentified

Empirics — Classic Bartik

$$gEmp_{jt}^{N} = \beta_0 + \underbrace{\beta_1}_{Short\ \ run\ \ response} m_{jt} + \eta_{jt}$$

Summary of Emprical Results

- Using Google Trends we propose a new instrument
- We find
  - In linguistically assimilated areas $\uparrow$ in immigration $\uparrow$ native employment

Results.

TABLE 3. Immigration and growth in log native employment in integrated enclaves

LHS = $\%\Delta Emp_N$	(1) OLS	(2) OLS	(3) IV	(4) IV
Immigration	0.001***	-0.000	0.007***	0.016**
	(0.000)	(0.000)	(0.001)	(0.007)
Constant	-0.001**	0.001	-0.007***	-0.014**
	(0.000)	(0.000)	(0.001)	(0.006)
Region FE	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes
First stage F stat			157.92	317.17
Mean LHS	0.00020	0.00020	0.00020	0.00020
N	896	896	896	896

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Panel data estimations. The dependent variable is growth in log native employment Time period for analysis 2013-2019. Shares are defined in 2011 using generic queries ir Google Trends. Specification (1) and (2) are OLS controlling for region or region and yea fixed effects respectively. Specifications (3) and (4) use integrated enclaves as instrument controlling for region or region and year fixed effects respectively. Data on employment stock of foreigners, and population come from Eurostat. Data on shares come from D4 project. Data on inflows come from OECD.

Results.

TABLE 4. Immigration and growth in log native employment in isolated enclaves

LHS = $\%\Delta Emp_N$	(1) OLS	(2) OLS	(3) IV	(4) IV
Immigration	0.001***	-0.000	0.010**	-0.004
	(0.000)	(0.000)	(0.004)	(0.006)
Constant	-0.001**	0.001	-0.010**	0.003
	(0.000)	(0.000)	(0.004)	(0.005)
Region FE	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes
First stage F stat			156.41	317.05
Mean LHS	0.00020	0.00020	0.00020	0.00020
N	896	896	896	896

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Panel data estimations. The dependent variable is growth in log native employment. Time period for analysis 2013-2019. Shares are defined in 2011 using generic queries in Google Trends. Specification (1) and (2) are OLS controlling for region or region and year fixed effects respectively. Specifications (3) and (4) use isolated enclaves as instruments controlling for region or region and year fixed effects respectively. Data on employment, stock of foreigners, and population come from Eurostat. Data on shares come from D4I project. Data on inflows come from OECD.

Results.

TABLE 5. Integrated enclaves with a different starting year

LHS = $\%\Delta Emp_N$	(1) OLS	(2) OLS	(3) IV	(4) IV
Immigration	0.001***	-0.001*	0.007***	0.016***
	(0.000)	(0.000)	(0.000)	(0.004)
Constant	-0.001*	0.001	-0.007***	-0.017***
	(0.000)	(0.000)	(0.000)	(0.004)
Region FE	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes
First stage F stat			131.53	280.73
Mean LHS	0.00028	0.00028	0.00028	0.00028
N	768	768	768	768

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Panel data estimations. The dependent variable is growth in log native employment. Time period for analysis 2014-2019. Shares are defined in 2011 using generic queries in Google Trends. Specification (1) and (2) are OLS controlling for region or region and year fixed effects respectively. Specifications (3) and (4) use integrated enclaves as instruments controlling for region or region and year fixed effects respectively. Data on employment, stock of foreigners, and population come from Eurostat. Data on shares come from D4I project. Data on inflows come from OECD.

Results.

TABLE 6. Isolated enclaves with a different starting year

LHS = $\%\Delta Emp_N$	(1) OLS	(2) OLS	(3) IV	(4) IV
Immigration	0.001***	-0.001*	0.005***	-0.019
	(0.000)	(0.000)	(0.002)	(0.073)
Constant	-0.001*	0.001	-0.005***	0.019
	(0.000)	(0.000)	(0.002)	(0.073)
Region FE	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes
First stage F stat			127.50	279.09
Mean LHS	0.00028	0.00028	0.00028	0.00028
N	768	768	768	768

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Panel data estimations. The dependent variable is growth in log native employment. Time period for analysis 2014-2019. Shares are defined in 2011 using generic queries in Google Trends. Specification (1) and (2) are OLS controlling for region or region and year fixed effects respectively. Specifications (3) and (4) use isolated enclaves as instruments controlling for region or region and year fixed effects respectively. Data on employment, stock of foreigners, and population come from Eurostat. Data on shares come from D4I project. Data on inflows come from OECD.

Results.

Appendix H: Descriptive Statistics and variables description

	N	mean	sd	min	max
isolation index[a]	896	16.2	15.7	1.40	110
isolation index[b]	896	11.8	10.3	1	48.0
isolation index[c]	896	13.9	12.7	1.22	70.5
isolation index[d]	896	16.3	14.2	1.52	76.6
Clustering	896	1.00	0.61	0	2.33
Population	896	3,055	2,676	349	10,116
Inflow	896	64,780	78,212	2,034	259,411
Change in foreign stock	896	23.9	29.5	-3.40	193
Change in supply of immigrants	896	0.78	0.57	-0.37	2.36
Initial integrated shares	896	0.063	0.098	0	0.46
Integrated shares[a]	896	0.063	0.064	0	0.22
Integrated shares[b]	896	0.063	0.062	0	0.21
Integrated shares[c]	896	0.063	0.066	0	0.25
Integrated shares[d]	896	0.063	0.066	0	0.24
Initial isolated shares	896	0.062	0.13	0	0.80
Isolated shares[a]	896	0.062	0.12	0	0.59
Isolated shares[b]	896	0.063	0.14	0	0.93
Isolated shares[c]	896	0.062	0.13	0	0.77
Isolated shares[d]	896	0.063	0.11	0	0.68
Daily frequency of internet usage	896	62.3	7.36	49.3	77.1
Weekly frequency of internet usage	896	76.2	6.11	65.2	92.0
Growth in log native employment	768	0.00020	0.0023	-0.0052	0.010

[a] Based on generic queries in 2011.
[b] Based on generic queries in January-February 2011.
[c] Based on generic queries in November-December 2011.
[d] Based on common queries in 2011.

Future work

Broader cultural connections

- Linguistic dissimilarities across Hispanic groups in the United States.

- Correlation of search topics between foreign country's population and immigrant enclaves in the United States.

Appendix

- queries queries
- occupations occupations
- model model
- literature review litrev
- NUTS2 geo
- data data
- language groups langGroups

1. Economic model
   - predicts positive and negative effects of immigration on native employment
   - positive effect - immigrants have higher search costs and lower wage which allows firms to open more vacancies
   - negative effect - replacement effect
2. Novel method to estimate linguistically assimilated enclaves
   - Using Google Search data
   - Areas where immigrants reside but don't search in their language in 2011 - linguistically assimilated
   - New inflows of immigrants to those areas boost native employment in contrast to linguistically unassimilated areas

Back to beginning

Dustmann and Van Soest 2002

Table 2. Cross-Tabulations, Speaking Fluency, Men.

t − 1/t	Very Poor		Poor		Intermediate		Good		Very Good		Total
Very Poor	**12**	**(26.67)**	21	(46.67)	8	(17.78)	3	(6.67)	1	(2.22)	46
Poor	16	(3.05)	**281**	**(53.63)**	190	(36.26)	35	(6.68)	2	(0.38)	524
Intermediate	4	(0.31)	186	(14.19)	**746**	**(56.90)**	343	(26.16)	32	(2.44)	1,311
Good	2	(0.15)	37	(2.86)	321	(24.79)	**787**	**(60.77)**	148	(11.43)	1,295
Very Good	1	(0.19)	3	(0.57)	34	(6.51)	186	(35.63)	**298**	**(57.09)**	522
Total	35	(0.95)	528	(14.28)	1,299	(35.14)	1,354	(36.62)	481	(13.01)	3,697

Numbers refer to 1984–1987. Column entries: previous year. Row entries: current year. Numbers in parentheses: Transition Probabilities, year $t-1$ to year t.

Back to beginning

alpha

$$g(f_\lambda(word))_j = 100 \times \frac{(w_k)}{(w_j)} \times \frac{25/200}{15/100}$$

$$g(f_\lambda(word))_j = 100 \times \frac{(w_k)}{(w_j)} \times \frac{25/200}{15/100}$$

$$\alpha_{\lambda j} = g(f_\lambda(word))_j = 100 \times \frac{(w_k)}{(w_j)} \times \frac{25/200}{15/100} \times \frac{200/100}{20/60} =$$

$$g(f_\lambda(word))_j = 100 \times \frac{(w_k)}{(w_j)} \times \frac{25/200}{15/100}$$

$$\alpha_{\lambda j} = g(f_\lambda(word))_j = 100 \times \frac{(w_k)}{(w_j)} \times \frac{25/200}{15/100} \times \frac{200/100}{20/60} =$$

$$= 100 \times \frac{(w_k)}{(w_j)} \times \frac{25/60}{15/20} \approx 53$$

Queries Back to back to graphs

Common queries

music
movies
news
songs
weather
restaurants
table
maps
map
translate
calculator
dictionary
games
postal code

entertainment
translator
flights
sports
directions
food
clothes
girlfriend
boyfriend
relationship
dating
t-shirts
umbrellas
pants
cars

browser
doctor
massage
recipes
time
exercises
credit cards
credit card
bank loan
tax
income tax
loans
loan
money
rent

housing
apartments
shoes
sneakers
business
lunch
breakfast
lunch
dinner
buy a house
sales
discounts
coupons
download

Common queries

bankruptcy
marriage
divorce
hospital
birth
funeral
wedding
church
mosque
temple
bus
airline
bus
schedule
a resume
jobs

job
vacancies
training
hire
wages
welfare
health
insurance
unemployment
school
education
university
diploma
public school
private
school
certificate

qualification
bachelor
degree
master's
degree
PhD degree
textbooks
books
career
salary
work
employment
find a job
labor
law

Generic queries

red
orange
yellow
green
blue
white
black
one
two
three
four
five
six
seven
eight
nine
ten
eleven
twelve
thirteen
fourteen
fifteen
sixteen
seventeen
eighteen
nineteen
twenty
thirty
forty
fifty
sixty
seventy
eighty
ninety
hundred
thousand
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
January
February
March
April
May
June
July
August
September
October
November
December
he
she
you
we
they

Generic queries

I
time
year
people
way
man
day
thing
child
government
work
life
woman
be
have

do
say
can
get
make
go
with
from
about
before
after
between
other
good
new

old
high
small
different
large
local
social
important
long
young

back to graphs

- Germany
  1. Legislators, senior officials, and managers 1 %
  2. Professionals 9 %
  3. Technicians and associate professionals 13%
  4. Clerks 7 %
  5. Service workers and shop and market sales workers 15 %
  6. Skilled agricultural and fishery workers 1 %
  7. Craft and related trades workers 20 %
  8. Plant and machine operators and assemblers 14 %
  9. Elementary occupations 20 %
  10. Armed forces 0 %