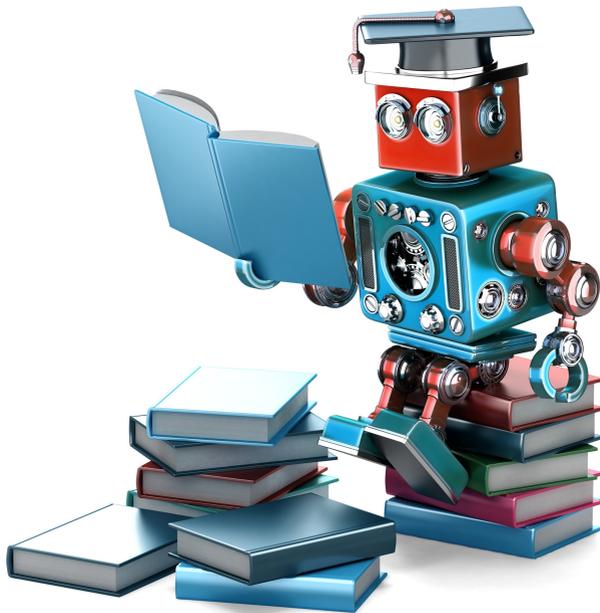




HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Computerlinguistik- Quiz

Heinrich-Heine-Universität Düsseldorf



studycl.phil.hhu.de

Ist Computerlinguistik der richtige Studiengang für mich?

Warum kann Google Translate keine Redewendungen übersetzen? Wieso klingt mein Navi so komisch, wenn ich im Ausland unterwegs bin? Und warum verschlimmbessert Autocorrect meine Nachrichten auf dem Smartphone immer auf unpassendste Weise?

Die Computerlinguistik ist eine Wissenschaft, die sich mit diesen und ähnlichen Fragen auseinandersetzt. Wir haben dieses Quiz zusammengestellt, um Studieninteressierten einen Einblick in den Alltag von Computerlinguist_innen zu geben.

Wenn du an den Quizaufgaben Spaß hast und die richtigen Antworten findest, wirst du im Studiengang Computerlinguistik gut aufgehoben sein! Auf unserer Webseite (studycl.phil.hhu.de) findest du weitere Informationen zum Studium.

Computerlinguistik-Projekte

Aufgaben wie die hier vorgestellten kann man alleine oder in Gruppen lösen. Im Computerlinguistik-Studium gibt es immer wieder Gelegenheiten, Projekte im Team zu bearbeiten. Im Jahrgang 2018/19 wurde z.B. ein Sprichwort-Generator als Gruppenprojekt im ersten Semester entwickelt:



Link zum Projekt: botsin.space/@python2018bot

1. SUPERWITZIGE UROMA

In Wörterbüchern kann man nachlesen, wie Wörter geschrieben werden. Aber Wörterbücher nehmen viel zu viel Platz weg! Selbst wenn sie online gespeichert sind, brauchen sie Festplattenplatz auf einem Server im Internet.

Zum Glück gibt es sparsamere Möglichkeiten, Wörter zu beschreiben, zum Beispiel mit regulären Ausdrücken. Mit ihnen kann man ganze Gruppen von Wörtern auf einmal beschreiben. Man muss nur wissen, was diese Wörter jeweils gemeinsam haben. Die Tabelle zeigt einige reguläre Ausdrücke und ihre Bedeutungen.

Findet ihr die regulären Ausdrücke, die in die letzten drei Zeilen gehören?

(UR)*OMA	Alle Wörter, die mit OMA enden und mit o oder beliebig vielen Wiederholungen von UR beginnen: OMA, UROMA, URUROMA, URURUROMA, ...
(PRAK ESS)TISCH	Ausschließlich Wörter, die entweder mit PRAK oder ESS beginnen und mit TISCH enden: PRAKTISCH, ESSTISCH
SU+PERWITZIG	Alle Wörter, die mit S beginnen und mit PERWITZIG enden und dazwischen entweder ein oder mehrere U haben: SUPERWITZIG, SUUPERWITZIG, SUUUUPERWITZIG, ...
...UNG	Alle Wörter, die mit UNG enden und davor genau drei Zeichen haben: ATMUNG, AHNUNG, EHRUNG, LESUNG, SPRUNG, ...
.*END?	Alle Wörter, die mit EN oder END enden (das D ist optional: es kann also da sein, muss aber nicht) und davor beliebig viele beliebige Zeichen haben: ESSEN, ESSEND, PROGRAMMIEREN, PROGRAMMIEREND, FRAGEN, FRAGEND, ...
(1)_____	Alle Wörter, die mit HURR beginnen und einem oder mehreren A plus einem optionalen H enden: HURRA, HURRAAA, HURRAH, HURRAAAAAH, ...
(2)_____	Ausschließlich die folgenden Wörter: ANWALT, INHALT, URALT
(3)_____	Alle Wörter, die mit SPRACH beginnen: SPRACHERWERB, SPRACHWISSENSCHAFT, SPRACHENTWICKLUNG, SPRACHENVIELFALT, ...

2. ES WERDE LICHT! ODER AUCH NICHT...

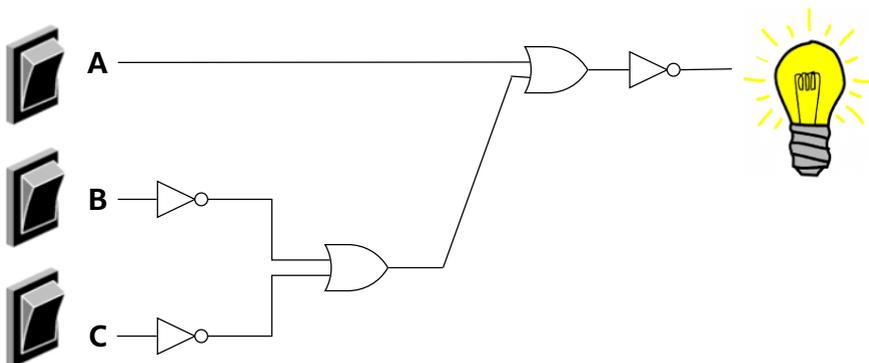
Die Stromleitungen im Sekretariat der Computerlinguistik sind etwas komplizierter konstruiert, als sie sein müssten. Die Bauteile ODER und NICHT wurden entsprechend der Zeichnung verbaut.

Der Strom fließt von den Schaltern links durch die Leitungen und Logikbausteine. Rechts muss Strom ankommen, damit die Lampe leuchtet.

Das Symbol $A \text{ --- } \neg \text{ --- out}$ steht für "NICHT". Dieser Baustein gibt genau dann Strom nach rechts weiter, wenn links kein Strom fließt. Wenn links Strom fließt, gibt der Baustein keinen Strom nach rechts weiter.

Das Symbol $\begin{matrix} A \\ B \end{matrix} \text{ --- } \cup \text{ --- out}$ steht für "ODER". Dieser Baustein gibt genau dann Strom nach rechts weiter, wenn durch eine oder beide Leitungen links Strom fließt (nur A, nur B, oder A und B gleichzeitig).

Welche Schalter müssen angeschaltet werden, damit die Lampe angeht?



Die richtigen Einstellungen könnt ihr unten ankreuzen. Die Buchstaben in der letzten Zeile braucht ihr später für das Lösungswort.

A		B		C	
AN	AUS	AN	AUS	AN	AUS
<input type="checkbox"/>					
S	C	H	N	U	R

3. REISEN, WEBEN, SINGEN

Samoa ist ein Inselstaat in Polynesien, in dem etwa 200.000 Menschen leben. Da Samoa sehr nah am Äquator liegt, ist es dort das ganze Jahr über tropisch heiß.

Die Amtssprachen in Samoa sind Englisch und Samoanisch. Englisch wird hier in Deutschland an allen Schulen unterrichtet, aber kann jemand von euch Samoanisch?



Die Tabelle zeigt Übersetzungen einiger Wörter aus der samoanischen Sprache. Wie werden die **fehlenden Wörter** korrekt übersetzt?

Tip: Die Wörter, die zusammen in einer Zeile stehen, haben immer etwas gemeinsam...

manao	sie wünscht	mananao	sie wünschen
matua	sie ist alt	matutua	sie sind alt
malosi	sie ist stark	malolosi	sie sind stark
punou	sie biegt	punonou	sie biegen
atamaki	sie ist weise	atamamaki	sie sind weise
savali	sie reist	(1) _____	sie reisen
laga	sie webt	(2) _____	sie weben
(3) _____	sie singt	pesese	sie singen

4. WIE MAN'S SPRICHT

IPA ist eine **Lautschrift**, mit der wir Wörter in verschiedenen Sprachen so aufschreiben können, wie sie ausgesprochen werden. Dabei repräsentiert ein Zeichen je einen Laut. Manchmal produzieren zwei oder drei Buchstaben zusammen einen Laut, müssen also mit einem einzigen IPA-Symbol übersetzt werden!

Die Tabelle zeigt einige Transkriptionen (Übersetzungen in IPA) mit den dazugehörigen Wörtern.

mich	[mɪç]	Buch	[buχ]
Bücher	[byçɐ]	Spruch	[ʃpʁuχ]
Gerücht	[gɛʁyçt]	hoch	[hoχ]
mechanisch	[mɛçanɪʃ]	doch	[dɔχ]
Recht	[ʁɛçt]	Bach	[baχ]
Löcher	[lœçɐ]	nach	[naχ]
leicht	[laɪçt]	rauchen	[ʁaʊχn]

Wie werden die drei Wörter unten transkribiert? Die Buchstaben an den richtigen Antworten braucht ihr später für das Lösungswort.

(1) Schacht	
[ʃaχt]	F
[ʃɛχt]	L
[ʃaçt]	U
[ʃɛçt]	G

(2) Becher	
[bɛχɛɐ]	B
[bɛçɛɐ]	L
[bɛχɐ]	A
[bɛçɐ]	U

(3) reichen	
[ʁaɪçn]	G
[ʁaɪχn]	I
[ɛaɪçn]	F
[ɛaɪχn]	T

LÖSUNGSWORT

Um zu gewinnen, müsst ihr das Lösungswort zusammensetzen.
Dazu braucht ihr ...

1. Das **zweite Zeichen** aus dem regulären Ausdruck Nummer (1).
Das **vorletzte Zeichen** aus dem regulären Ausdruck Nummer (2).
Das **vierte Zeichen** aus dem regulären Ausdruck Nummer (3).
2. Die drei Buchstaben, die ihr bei Aufgabe 2 angekreuzt habt.
3. Den **ersten Buchstaben** aus der Übersetzung Nummer (1).
Den **ersten Buchstaben** aus der Übersetzung Nummer (2).
Den **dritten Buchstaben** aus der Übersetzung Nummer (3).
4. Die drei Buchstaben, die ihr bei Aufgabe 4 angekreuzt habt.

Ach ja... und das Lösungswort ist leider etwas durcheinandergeraten.
Könnt ihr die Buchstaben so umstellen, dass das richtige Wort herauskommt?

Lösungswort:

LÖSUNGEN UND LÖSUNGSANSÄTZE

1. Superwitzige Uroma

(1) Mit einem + kann man beliebig viele (aber mindestens ein) Zeichen generieren. Ein ? bedeutet ein optionales Vorkommen eines Zeichens oder einer Zeichengruppe.

Lösung: **HURRA+H?**

(2) Die drei Wörter **ANWALT**, **INHALT**, **URALT** haben die gemeinsame Endung **ALT** und unterscheiden sich nur durch die Anfangsbuchstaben.

Lösung: **(ANW|INH|UR)ALT**

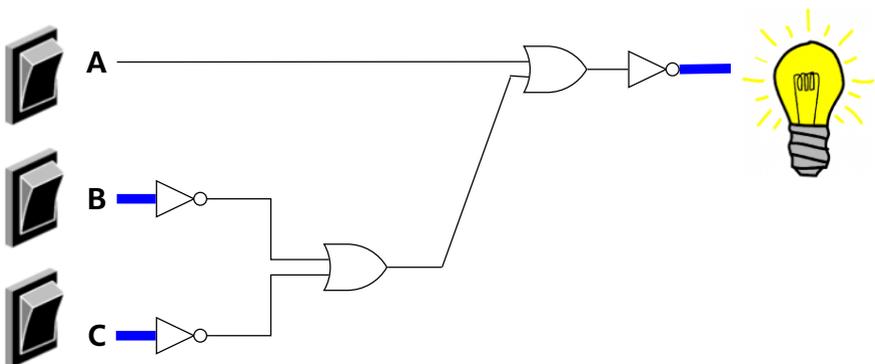
(3) Die Wörter, die mit **SPRACH** beginnen und beliebig viele beliebige Zeichen danach haben.

Lösung: **SPRACH.***

2. Es werde Licht! Oder auch nicht...

Damit die Lampe angeht, muss der Schalter **A** ausgeschaltet sein, aber die beiden Schalter **B** und **C** müssen an sein. Nur in diesem Fall kommt der Strom von allen Schaltern zur Lampe.

Die **blau** markierten Linien unten auf dem Bild zeigen den Stromfluss im Netz.



3. Reisen, weben, singen

Aus den Wortformen in der Tabelle kann man Regelmäßigkeiten in der Wortbildung der samoanischen Sprache ableiten. Offensichtlich wird die 3. Person Pluralform der Verben im Samoanischen durch die Wiederholung der zweiten Silbe gebildet.

Daher lauten die Lösungen folgendermaßen:

- (1) **savavali** — sie reisen
- (2) **lagaga** — sie weben
- (3) **pese** — sie singt

4. Wie man's spricht

(1) Der erste Laut im Wort **SCHACHT** wird mit [ʃ] transkribiert. Der zweite Laut wird wie "a" im Wort "*mechanisch*" ausgesprochen, und daher mit [a] transkribiert. **CH** wird in diesem Wort wie in "*Buch*" ausgesprochen, und daher mit [x] wiedergegeben. Der Laut **T** wird mit [t] transkribiert.

Daher lautet die Lösung [ʃaxt].

(2) Analog zum Lösungsweg in der Teilaufgabe (1) kann man feststellen, dass die korrekte Transkription für das Wort **BECHER** — [bɛçɐ] ist.

(3) Die richtige Transkription für das Wort **REICHEN** lautet [ʀaɪçn].

Lösungswort

U-L-A-C-H-U-S-L-S-F-U-G



S-C-H-U-L-A-U-S-F-L-U-G

COMPUTERLINGUISTIK IN DÜSSELDORF STUDIEREN

Computerlinguistik = Linguistik + Informatik

Die Computerlinguistik ist ein etabliertes Schnittstellenfach, das Methoden der Informatik auf Fragestellungen der Linguistik anwendet. Auf der theoretischen Seite wird versucht, menschliche Sprache mittels mathematisch fundierter Modelle zu analysieren und zu verstehen. Diese Modelle werden auf der praktischen Seite dazu verwendet, Softwarewerkzeuge zu entwickeln, die eine maschinelle Verarbeitung natürlicher Sprachdaten ermöglichen.



Studienaufbau Bachelor Computerlinguistik

Im Bachelor Computerlinguistik werden sowohl theoretische als auch praktische Aspekte des Fachs vermittelt. Wer sich für Sprache interessiert, aber auch den Disziplinen Informatik und Mathematik gegenüber aufgeschlossen ist, ist in diesem Studiengang gut aufgehoben. Programmierkenntnisse sind keine Voraussetzung und werden im Rahmen der Grundlagenmodule praxisbezogen vermittelt. Nach den Basismodulen erfolgt eine Spezialisierung in informatischen und linguistischen Bereichen.

Karriereaussichten

Da der computerlinguistische Sektor in Industrie und Wirtschaft an Bedeutung gewinnt, ist der Bedarf an gut ausgebildeten Computerlinguist_innen in letzter Zeit stark gestiegen. Sie können zum Beispiel Aufgaben in den folgenden Tätigkeitsfeldern wahrnehmen:

- Maschinelle Übersetzung;
- Textgenerierung. z.B. für Onlineshops;
- Sprachdatenanalyse zu Marketingzwecken (Social Media Monitoring);
- Lexikographische Tätigkeiten (computergestützte Wörterbucharstellung);
- Entwicklung oder Optimierung von Suchmaschinen;
- Entwicklung von Dialogsystemen wie z.B. Siri, Cortana, Google Now.

Ansprechpartnerinnen

Prof. Dr. Laura Kallmeyer

E-Mail: kallmeyer@phil.uni-duesseldorf.de

Prof. Dr. Wiebke Petersen

E-Mail: petersen@phil.uni-duesseldorf.de