

# Fotis Jannidis / Hubertus Kohle/ Malte Rehbein (Hg.): *Digital Humanities. Eine Einführung*

## Übungsaufgaben

### I.1 Geschichte der Digital Humanities

- Vergleichen Sie das Inhaltsverzeichnis des ersten Bandes der Zeitschrift *Computers and the Humanities* (<http://dblp2.uni-trier.de/db/journals/lre/lre1.html>) mit den Themen die im Abstractband der Jahreskonferenz der *digital humanities im deutschsprachigen Raum* von 2016 behandelt werden (<http://dhd2016.de/boa.pdf>). Welche Themen sind geblieben, verschwunden, dazu gekommen?
- Was wird aus den beschriebenen Trends in Ihrer inhaltlichen Disziplin auf dem Smartphone?
- Wenn Sie heute Zugang zu einer der industriellen Führungsfiguren des Silicon Valley hätten: Zur Finanzierung der Anwendung welcher brandneuen Technologie auf Ihr inhaltliches Fach würden Sie sie gerne überreden?

### I.2 Digital Humanities als Wissenschaft

- Was an der „Knochenarbeit“ in Ihrem inhaltlichen Fach halten Sie für automatisierbar?
- Welche Form der Beweisführung in Ihrem inhaltlichen Fach könnte von einer empirischen Absicherung profitieren?
- Können Sie sich eine „Simulation“, wie im 3D-Bereich beschrieben, für ein anderes Medium Ihrer Wahl vorstellen? (Beispiel: Simulation von Klanglandschaften).
- Schlagen Sie die Definitionen von „Syntax“ und „Semantik“ nach. Die meisten textuellen Beispiele haben klare Bezüge zum syntaktischen Bereich. Wieweit sind semantische Bezüge auf dem Rechner darstellbar?

### I.3 Aufbau des Computers und Vernetzung

- Erklären Sie, wie eine einfache Rechenaufgabe in Schaltalgebra überführt wird.
- Nennen Sie drei Gründe, die für das Schichtenmodell sprechen.
- Berechnen Sie den Speicherplatzbedarf eines Bildes auf einem Display in der Full HD-Auflösung von 1920x1080 Pixeln und einer Farbtiefe von 48 bit.
- Welche Übertragungsgeschwindigkeit kann mit einem „Sneakernet“, also dem Transport von Datenträgern von einem Punkt zum anderen, erreicht werden und wann ist dies sinnvoll?
- Welche Chancen und welche Gefahren liegen im Cloud-Computing?

- Sind Sie der Meinung, dass auch IPv6-Adressen irgendwann gering verfügbar sein könnten?
- Nennen Sie alle Kontakte mit Computernetzen in Ihrem Alltag und beschreiben Sie, was passieren würde, wenn diese nicht mehr funktionierten.

## I.5 Zahlen und Zeichen

- Wandeln Sie die Dezimalzahl 27 in eine Binärzahl um.
- Wandeln Sie die Binärzahl 101011001 in eine Dezimalzahl um.
- Welcher Dezimalzahl entspricht die Zahl FFh?
- Für welche Standards ist das W3C-Konsortium zuständig?
- Schlagen Sie den Codepoint und den Unicode-Namen des Buchstabens Ö in der Unicode-Spezifikation nach.
- Finden Sie heraus, wie man feststellen kann, dass eine Datei in UTF-8 kodiert ist.

## II.7 Datenmodellierung

- Recherchieren Sie, welche Metadaten Ihre Bibliothek im Opac zu einem Titel anzeigt und welche weiteren für die Anzeige der Suchergebnisse evtl. noch vorhanden sein müssen.
- Recherchieren Sie in der Wikipedia, was der Unterschied zwischen Gleitkommazahlen und Festkommazahlen ist.
- Welche anderen Normdateien können Sie an der Deutschen Nationalbibliothek oder der Library of Congress finden?
- Arbeiten Sie in Jannidis/Flanders 2017 den Abschnitt 3: “Foundations“ durch.

## II.8 Datenbanken

- Was ist Sinn und Vorteil der Drei-Ebenen-Architektur?
- Was sind die zentralen Begriffe des Entity-Relationship-Modells?
- Wie ist aus dem konzeptuellen Modell einer Datenbank ein logischer Entwurf abzuleiten?
- Welches sind die Schritte der Normalisierung und wozu sind sie notwendig?
- Welches sind die Grundelemente einer SQL-Abfrage?

## II.9 XML (Lösungen zu den Aufgaben zu diesem Kapitel finden Sie unten)

### 1. Betrachten Sie diesen XML-Schnipsel:

```
<satz autor="Schiller, Friedrich">Wo der
<alternativ><original>Brodgelehrte</original>
<modernisiert>Brotgelehrte</modernisiert></alternativ> trennt, vereinigt der
philosophische Geist.</satz>
```

- a. Welche Teile davon sind Code bzw. Auszeichnungen?
  - b. Was sind hier tag, Element, Attribut, Attributwert und Elementinhalt (zu welchem Element)?
2. a. Wo beginnt im folgenden Beispiel der Baum der XML-Struktur?
- ```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/css" href="anzeige.css"?>
<?xml-model href="schiller.rng"?>
<satz autor="Schiller, Friedrich">Wo der
<alternativ><original>Brodgelehrte</original><modernisiert>Brotgelehrte</mod
ernisiert></alternativ> trennt, vereinigt der philosophische Geist.</satz>
```
- b. Zeichnen Sie die Baumstruktur des XML-Textes auf.
  - c. Wie heißt das Wurzelement im Beispiel?
  - d. Wie heißt das erste Kindelement des Elements `alternativ`?
  - e. In welchem Verhältnis stehen, die Elemente `original` und `alternativ` zu einander?
3. a. Was versteht man unter „prozeduraler“, „präsentationsorientierter“ und „deskriptiver“ Textauszeichnung?
- b. Wie unterscheiden sich eine „präskriptive“ und eine „deskriptive“ Verwendung von XML? In welchen Situationen würden Sie die beiden Methoden jeweils anwenden?
- c. Was ist mit „mixed content“, gemeint?
4. a. Für welche Zwecke können Sie ein (eigenes oder ein von anderen definiertes) Schema in einem XML-Dokument verwenden?
- b. Welche der folgenden Eigenschaften eines XML-Dokuments kann man mit einem Schema festlegen: Elementnamen, Inhalte von Attributen, Herkunft der Daten eines XML-Dokuments, Reihenfolge von Elementen, Inhaltsstrukturen von Elementen, das Entstehungsdatum eines Textes?
5. a. Gegeben ist der XPath-Ausdruck `distinct-values(//vers/*/name())`
- (a) Drücken Sie den XPath-Ausdruck auf Deutsch aus.
  - (b) Was wird zurückkommen?
  - (c) Welcher Art ist die Rückgabe?
- b. Formalisieren Sie und übersetzen Sie auf XPath: Welche Elemente sind mit einer GND-Identifikation ausgestattet? (Hinweis: Sie müssen eine Bedingung formulieren).

6. a. Wozu dient XSLT?
- b. Welche Formate kann der Output eines XSLT-Prozessors haben?
- c. Welche der folgenden XSL-Befehle ist eine XSLT-Schablone (*template*) für Verse?

```
<xsl:value-of select="vers" />
<xsl:template match="text" />
<xsl:template match="vers" />
<xsl:apply-templates select="vers" />
```

## II.10 Netzwerke

- Verwenden Sie Gephi, um den Graphen aus Abb. 7 zu visualisieren.
- Bei einer Party, bei der sieben Männer und sechs Frauen anwesend sind, begrüßen alle Frauen alle Männer, zwei Drittel der Frauen begrüßen sich untereinander und drei der Männer begrüßen sich untereinander, wobei auffällt, dass es bei Männern und Frauen nur Grüßer und Nicht-Grüßer gibt. Grüßer begrüßen alle anderen Grüßer (und alle des anderen Geschlechts), während Nicht-Grüßer nur Vertreter des anderen Geschlechts grüßen. Wie sieht das Netzwerk aus? Wieviele Kanten hat es insgesamt. Was sind der maximale und der minimale Grad des Netzwerks?
- Berechnen Sie *betweenness centrality* für den Graphen in Abb. 7. Sie können dazu jede beliebige Software oder Bibliothek verwenden.
- Lesen Sie den Beitrag von Apoorv Agarwal: Social Network Analysis of Alice in Wonderland. In: *Workshop on Computational Linguistics for Literature* Montreal, Canada, June 8, 2012, S. 88–96. Diskutieren Sie, welche Probleme der Ansatz nach Aussage der Autoren hat und welche anderen Probleme Sie sehen.

## II.11 Ontologien

- Ergänzen Sie das Instanzen-Modell von Abb. 3 um die Instanzen Heinrich Mann und Theodor Mommsen, wie sie im Text charakterisiert sind.
- Vervollständigen Sie die RDF/Turtle-Notation unserer Beispiel-Ontologie. Modellieren Sie anschließend zusätzlich die Geburtsorte der Personen, sowie den Vergabeort der Nobelpreise.
- In unserer Beispiel-Ontologie haben wir z.B. den Literaturnobelpreis als eine Instanz der Klasse Nobelpreis modelliert. Eine Alternative wäre gewesen, ihn als Unterklasse zu Nobelpreis aufzufassen. Diskutieren Sie die Unterschiede. Ergänzen Sie die Ontologie um das Jahr des Preisgewinns. Beachten Sie dabei die Nuancen der beiden Aussagen: „Theodor Mommsen ist Literaturnobelpreisträger“ und „Theodor Mommsen war Literaturnobelpreisträger von 1902“.
- Informieren Sie sich an Hand von Lexika über die Debatte zum Geburtsdatum Karls des Großen. Modellieren Sie die verschiedenen Sichtweisen mittels RDF-Reification.
- (umfangreich): Modellieren Sie die Beispielwelt der Nobelpreise mit Hilfe von CIDOC CRM. Berücksichtigen Sie dabei auch, in welchem Jahr die Preise gewonnen wurden.

### III.12 Digitalisierung

- Eine Tageszeitung soll seitenweise mit 300ppi und einer Farbtiefe von 16 Bit digitalisiert werden. Berechnen Sie die Bildgröße für ein einzelnes Digitalisat sowie den Speicherbedarf für einen ganzen Jahrgang.
- Mikroformen erlauben verkleinerte Abbildungen von bis zu 1/40 der Originalgröße (Saffady 2000, S. 4). Was bedeutet dies für die Auflösung bei der Digitalisierung solcher Mikroformen?
- Stellen sie Mikroverfilmung, Digitalisat und Original am Beispiel einer historischen Zeitung gegenüber: welche Vor- und Nachteile sehen Sie in Bezug auf Verbreitung und Zugänglichkeit, Nutzung durch die Forschung, Speicherung und Langzeitsicherung?
- Da auch Metadaten Modelle darstellen, gelten auch bei ihnen die Prinzipien von Reduktion, Subjektivität und Zweckgebundenheit. Erläutern Sie dies am Beispiel der Digitalisierung einer einem Archiv hinterlassenen handschriftlichen Briefsammlung und legen Sie dar, welche Auswirkungen dies bei der Suche in einer Datenbank hat.

### III.13 Digitales Publizieren

- Lesen Sie den Artikel von Mons/Velterop und finden Sie Argumente dafür und dagegen, dass sich die Nano-Publikation auch für den geisteswissenschaftlichen Bereich eignet.
- Finden Sie neben den Genannten weitere Gründe für die Vorzüge und Nachteile des Open Access.
- Zeigen Sie weitere Bereiche, in denen die Digitalisierung die „Granularisierung der Gesellschaft“ forciert.
- Vergleichen Sie drei selbstgewählte biographische Wikipedia-Artikel mit den Einträgen in der *Encyclopedia Britannica* und bewerten Sie deren Qualität.

### III.14 Digitale Wissensproduktion

- Recherchieren Sie zu den verschiedenen Werkzeugen für kollaboratives Schreiben (siehe Kasten S. 210). Worin unterscheiden sich diese Werkzeuge? Für welche Art von Schreibprojekt eignet sich Ihrer Meinung nach welches Werkzeug besonders gut und aus welchen Gründen?
- Schlagen Sie bei Wikipedia einen Artikel zu einem Thema nach, mit dem Sie sich etwas auskennen. Welche Hinweise liefern die Diskussionsseite und die Versionsgeschichte auf den Entstehungsprozess des Artikels? Welche inhaltlichen Stärken und Schwächen erkennen Sie? Melden Sie sich bei Wikipedia an und verbessern Sie den Artikel.
- Auch die Kapitel 13 „Digitales Publizieren“, 20 „Quantitative Analyse“, 22 Digitale Rekonstruktion und Simulation“ und 23 „Informationsvisualisierung“ befassen sich mit spezifisch digitalen Methoden des Umgangs mit geisteswissenschaftlichen Daten. Lesen Sie eines dieser Kapitel und notieren Sie, inwiefern sich hier die Produktion

wissenschaftlicher Erkenntnis gegenüber etablierten, analogen Analyseverfahren verändert.

### **III.15 Bibliothek, Archiv, Museum**

- Recherchieren Sie, wie viele Bibliotheken (öffentliche und wissenschaftliche) es in Deutschland gibt, wie groß der jeweilige Medienbestand ist und wie viele Entleihungen es in einigen ausgewählten Bibliotheken gibt.
- Neben dem Bundesarchiv gibt es zahlreiche Landesarchive. Welche zusätzlichen Aufgaben haben diese Landesarchive in Ergänzung zum Bundesarchiv?
- Was versteht man unter Museologie und wo kann diese Disziplin an Hochschulen studiert werden?
- Suchen Sie in der Deutschen Digitalen Bibliothek nach dem Begriff „Clavichord“ und finden Sie heraus, um was es sich handelt und wo überall Sie es sich vor Ort anschauen können.
- Gehen Sie zur Bilddatenbank des Bundesarchivs und nutzen dort die Themensuche (z.B. Bild / Kunst / Malerei und Grafik / Europa). Wie viele Treffer werden angezeigt? Wie ist die Fotografie erschlossen bzw. welche Metadaten werden dafür verwendet? Können Sie Fotografien herunterladen, unter welchen Bedingungen?
- Welche Trefferzahl ergibt sich bei einer Suche in der „Hathi Trust’s digital library“, wenn nach dem Titel „Die Leiden des jungen Werther“ in deutscher Sprache mit kostenfreiem Zugang zum Volltext gesucht wird? Verändert sich die Trefferanzahl, wenn bei Autor noch der Begriff „Goethe“ hinzugefügt wird?

### **III.16 Aufbau von Datensammlungen**

- Wählen Sie eine der im Kasten „Beispiele für Datensammlungen“ genannten Sammlungen aus und finden Sie möglichst viel über die angebotenen Daten sowie den Prozess und die Verantwortlichkeiten beim Erstellen dieser Datensammlung heraus. Wo genau findet man die entsprechenden Informationen? Welche Informationen sind verfügbar, welche fehlen?
- Welche Unterschiede ergeben sich für den Aufbau von Datensammlungen zwischen verschiedenen Typen von Gegenständen? Vergleichen Sie den Aufbau von Textsammlungen mit dem Aufbau von Datensammlungen, die sich auf Bilder, Urkunden oder Noten beziehen. Welche Gemeinsamkeiten und welche Unterschiede ergeben sich für die sechs genannten Schritte?

### **IV.18 Manuelle und automatische Annotation**

- Welche Unterschiede lassen sich zwischen händischen und automatischen Annotierungsverfahren benennen? Welche Vor- und Nachteile sehen Sie bei den jeweiligen Verfahren?
- Was kennzeichnet die Philosophie und die Strukturierung und Organisation von TEI-Tagsets?

- Was bedeuten *precision* und *recall* im Kontext automatischer Verfahren und wie lassen sie sich überprüfen?
- Welche Vorteile bietet die Idee des Semantic Web?
- Sie haben Interesse an Raumkonzepten in literarischen Erzählungen und dafür ein Korpus von 100 Texten, das von einem Mitarbeiterteam bearbeitet werden kann, aufgebaut. Wie würden Sie konkret vorgehen, um ein Annotationsverfahren zu entwickeln?
- Auf TEI by example (<http://tei.byexample.org/>) finden Sie Anwendungsbeispiele für TEI-Markup anhand bekannter Texte der Weltliteratur. Entwickeln Sie eine TEI-Auszeichnung für einen Text Ihrer Wahl und vergleichen Sie die unterschiedlichen Lösungen.

#### IV.19 Information Retrieval

- Was ist die Aufgabe eines Information Retrieval Systems auf der einen Seite gegenüber den Daten und auf der anderen Seite gegenüber dem Benutzer?
- Was ist die Aufgabe eines Webcrawlers und wie viele Seiten müsste dieser pro Sekunde verarbeiten, um jede Seite des World Wide Webs einmal im Monat zu prüfen?
- Welche Probleme können in einem Suchindex entstehen, der keine Synonyme und Homonyme prüft?
- Welche Probleme können mit der einfachen Anwendung von PageRank entstehen und wie reagiert Google in der Praxis darauf?
- Welche Formen von Benutzerschnittstellen sind denkbar, die die Suchergebnisse nicht in Listen darstellen?

#### IV.20 Quantitative Analyse

- Finden Sie im Bestand des Oxford Text Archives – <http://ota.ox.ac.uk> – die beiden Erzählungen „The Scarlet Letter“ und „The House of the Seven Gables“ von Nathaniel Hawthorne und speichern Sie die plain text-Fassung auf ihrem Computer ab. Laden Sie beide Texte zugleich in die Voyant Tools – <http://voyant-tools.org/> – und finden Sie heraus, wie sich die absoluten und relativen Häufigkeiten des Begriffs „time“ in den beiden Texten zueinander verhalten. Explorieren Sie weitere Begriffe.
- Installieren Sie die Statistik-Umgebung R – <https://cran.r-project.org/> – und das Paket „stylo“ für R: <https://sites.google.com/site/computationalstylistics/stylo>. Stellen Sie sich mit Daten aus TextGrids Digitaler Bibliothek – <https://textgridrep.org/repository.html> – eine Sammlung von je vier Werken drei unterschiedlicher Autor/innen zusammen. Lernen Sie mit Hilfe des „Package stylo“-Handbuchs, wie man „stylo“ verwendet und finden Sie heraus, welche der drei gewählten Autoren den ähnlichsten Stil haben.

## IV.22 Digitale Rekonstruktion und Simulation

- Lesen Sie die Einleitung von Nerdinger in dem von ihm herausgegebenen Ausstellungskatalog zur Geschichte der Rekonstruktion (2010) und markieren Sie seine sehr pointierte Position.
- Welches sind die entscheidenden Vorteile des digital gestützten Rekonstruierens gegenüber traditionell analogen Verfahren?
- Was empfinden klassische Historiker an der Simulation als verstörend?
- Welcher der bei Düring diskutierten drei Fälle scheint Ihnen der plausibelste und warum?

## V.24 Recht

- „Ein eBook ist kein Buch!“ – Begründen Sie diese Aussage mit juristischen Argumenten.
- Das Urheberrecht erlischt 70 Jahre nach dem Tod des Autors. Bedeutet das, dass in jedem Fall in Büchern oder Werkausgaben publizierte Texte von Immanuel Kant digitalisiert und ins Netz gestellt werden dürfen?
- In einer Datenbank werden Scans und transkribierte Volltexte eines Autors des 19. Jahrhunderts angeboten, die im letzten Jahr in handschriftlicher Fassung in seinem Nachlass gefunden wurden und bislang noch nicht publiziert worden sind. Sind die Datenbank und ihre einzelnen Elemente rechtlich geschützt?
- Erläutern Sie die Funktion von Creative-Commons-Lizenzen.

## V.25 Ethik

- Informieren Sie sich in einem Philosophie-Lexikon über die vier im Text erwähnten Schulen der Ethik. Recherchieren Sie, zu welchen Schulen folgende Prinzipien gehören: Kategorischer Imperativ, die rechte Mitte, größtes Glück der größten Zahl, Naturzustand.
- Lesen Sie das erste Kapitel in Hans Jonas' klassischer Schrift *Das Prinzip Verantwortung* (1979, S. 15-58). Inwieweit sind Jonas' Überlegungen auf die moralischen Probleme der Digital Humanities anwendbar und inwieweit nicht?
- Reflektieren Sie über Ihr eigenes Forschungs- oder Studienfeld. Wo ergeben sich möglicherweise moralische Probleme? Welche Verantwortlichkeiten ergeben sich für Ihre Person?



## Lösungen zu Kapitel II.9: XML

### Lösung 1.a

Die spitzen Klammern und alles von ihnen eingeklammerte, also z.B. `<satz autor="Schiller, Friedrich">`, `<alternativ>`, `<original>`, `</original>`, `</alternativ>`

### Lösung 1.b

**tag:** `<satz>`, `</satz>`, `<alternativ>`, `</alternativ>`, `<original>`, `</original>`, `<modernisiert>`, `</modernisiert>`

**Element:** `<satz autor="Schiller, Friedrich">Wo der <alternativ>`  
`<original>Brodgelehrte</original><modernisiert>Brotgelehrte</modernisiert>`  
`</alternativ> trennt, vereinigt der philosophische Geist.</satz>`  
`<alternativ><original>Brodgelehrte</original><modernisiert>Brotgelehrte</modernisie`  
`rt</alternativ>`  
`<original>Brodgelehrte</original>`  
`<modernisiert>Brotgelehrte</modernisiert>`

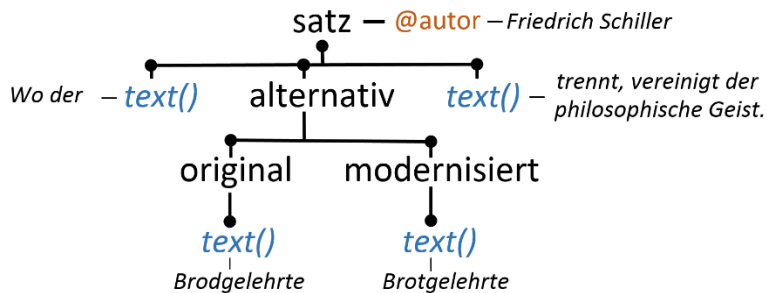
**Attribut:** `autor="Schiller, Friedrich"`; Attributwert: Schiller, Friedrich

**Elementinhalt:** Element `satz`: "Wo der Brodgelehrte Brotgelehrte trennt, vereinigt der philosophische Geist.", Element `alternativ`: "BrodgelehrteBrotgelehrte" Element `original`: "Brodgelehrte", Element `modernisiert`: "Brotgelehrte"

### Lösung 2.a

beim Element `satz`

### Lösung 2.b



### Lösung 2.c

`satz`

### Lösung 2.d

`original`

### Lösung 2.e

Sie sind Geschwister (*siblings*).

### **Lösung 3.a**

Prozedurales Markup rückt die Verarbeitung von Text in den Vordergrund, präsentationsorientiertes Markup ihre Darstellung, deskriptives Markup die Bedeutung von Textpassagen oder Informationsobjekten.

### **Lösung 3.b**

Wenn bei gegebenen Daten ein beschreibendes und strukturierendes Modell allmählich entwickelt wird, spricht man von einem deskriptiven Modell. Arbeitet man hingegen mit einem bereits bestehenden und durch ein Schema kodifizierten Modell, auf das Daten übertragen werden, nennt man dies präskriptiv. Der deskriptive Ansatz ist eher explorativ, heuristisch und für die Entwicklung neuer Modelle geeignet, während der präskriptive Ansatz für gleichmäßige, gut zu verarbeitende Daten sorgen kann. Im Lebenszyklus von Auszeichnungsvorhaben können sich beide Ansätze abwechseln und ergänzen.

### **Lösung 3.c**

XML-Elemente, die sowohl Text als auch andere Elemente enthalten.

### **Lösung 4.a**

Konsistenter Gebrauch von Namen von Elementen und Attributen; konsistente Verschachtelung von Elementen; Information über die angewendete Kodierung bei Datenaustausch; Erleichterung der Weiterverarbeitung eines XML-Dokuments; Formalisierung des im XML Dokument verwendeten Modells; u.v.a.m.

### **Lösung 4.b**

Elementnamen, Inhalte von Attributen, Reihenfolge von Elementen, Inhaltsstrukturen von Elementen

### **Lösung 5.a**

- (a) Wie heißen die verschiedenen Elemente, die in den Versen verwendet werden?
- (b) name redewendung
- (c) Sequenz von Strings.

### **Lösung 5.b:**

```
//*[contains(@identifikation,'gnd_')]
```

### **Lösung 6.a**

Zur Umwandlung von XML-Dokumenten in andere Dokumente.

## Lösung 6.b

direkt: alle text-basierten Formate (plain text; XML und damit auch XHTML, HTML, CSV, SVG; JSON), indirekt via XSL-FO auch PDF

## Lösung 6.c

```
<xsl:template match="vers" />
```