Inhaltsverzeichnis

$_{\rm I}$S$^{\rm I}$TEX
Interaktive Silbentrennung für die deutsche Sprache unter TEX 3.14 und 3.141 unter UNIX†

W. Barth H. Steiner H. Herbeck

23. August 1993

Zusammenfassung

Das speziell für deutschsprachige Texte entwickelte *SiSiSi*

SICHERE
SINNENTSPRECHENDE
SILBENTRENNUNG

kann jetzt in TEX und LATEX benutzt werden. Es gibt für UNIX change–files und alle anderen notwendigen Files auf dem anonymous–FTP–server `eiunix.tuwien.ac.at`. Dort findet man auch eine Installationsanleitung, mit der man das System leicht installieren kann. Dadurch wird ein neues System erzeugt, das das neue Verfahren für die Silbentrennung verwendet. Alle anderen TEX–Funktionen bleiben unverändert.

Es gibt zwei Varianten: Das einfache S$^{\rm I}$TEX und das interaktive, voll ausgebaute $_{\rm I}$S$^{\rm I}$TEX. Das einfache S$^{\rm I}$TEX kann nach der Installation genauso wie das ursprüngliche TEX verwendet werden. Es hat aber den Vorteil, daß es auch bei unerwarteten Wortzusammensetzungen keine falschen Trennstellen erzeugt. Welche zusätzlichen Möglichkeiten $_{\rm I}$S$^{\rm I}$TEX bietet, ist in diesem Bericht beschrieben.

† UNIX ist Markenzeichen von AT&T

1 Terminologie

TEX ist Donald E. Knuths Formatierungsprogramm für Texte [1], [2].

LATEX ist das von Leslie Lamport entwickelte Makro–Paket zu TEX, das im deutschsprachigen Raum meist verwendet wird [3], [4]. Ab sofort gilt TEX synonym für TEX und LATEX, und die im folgenden beschriebenen Erweiterungen SITEX und ISITEX können sowohl auf TEX als auch auf LATEX angewendet werden.

SITEX ist eine Variante von TEX für deutschsprachige Texte, die durch Austausch der „hyphenation“ durch das von W. Barth und H. Nirschl entwickelte Verfahren zur „sicheren sinnentsprechenden Silbentrennung (*SiSiSi*)“ für die deutsche Sprache [5], [6] entstanden ist.

ISITEX ist eine um einige zusätzliche Möglichkeiten, insbesondere interaktive Komponenten, erweiterte Form von SITEX.

2 SiSiSi

In der deutschen Sprache verwendet man gern lange, zusammengesetzte Wörter, wie z.B.:

Text=verarbeitungs=system oder Silben=trennungs=verfahren.

Wenn solche riesigen Gebilde nicht am Ende einer Zeile abgeteilt werden können, entsteht meist ein sehr unschönes Schriftbild. Das Hauptproblem bei der Silbentrennung ist das Finden der Nahtstellen (durch „=“ angezeigt) zwischen den Einzelwörtern. Die Trennstellen innerhalb der Einzelwörter lassen sich relativ leicht mit den im Duden [9] angegebenen Regeln ermitteln. *SiSiSi* findet die Zerlegung der Wörter mit Hilfe einer Worttabelle, genauer einer Tabelle aller Wortbestandteile: Vorsilben, Stämme und Endungen. Weniger als 8000 Eintragungen reichen aus, um fast alle deutschen Wörter und die gängigen Fremdwörter zu erfassen. Durch die Abstützung des Verfahrens auf eine Worttabelle ist es natürlich möglich, auch mit unvorhersehbaren, ungewöhnlichen Wortungetümen fertig zu werden. Dieses Zerspalten zusammengesetzter Wörter mit einer einfachen, überschaubaren

Methode bringt wesentliche Vorteile gegenüber dem aus dem Amerikanischen übernommenen pattern–Verfahren.

SiSiSi sucht nach allen möglichen Zerlegungen eines Wortes. Bei manchen gibt es mehr als eine, wie z.B.:

Bau=mast/Baum=ast oder Stau=becken/Staub=ecken.

Wir bezeichnen das als **Mehrdeutigkeit**. Stößt *SiSiSi* auf eine solche Mehrdeutigkeit, so nutzt die einfachere Implementierung SITEX die zweifelhafte Trennstelle nicht aus. Dadurch wird sichergestellt, daß keine falschen Trennungen entstehen. Das interaktive ISITEX hält während der Bearbeitung des Dokuments bei jedem solchen mehrdeutigen Wort an, zeigt dem Benutzer alle möglichen Zerlegungen und erlaubt ihm, die passende auszuwählen.

Falls *SiSiSi* für ein Wort keine Zerlegung findet, dann handelt es sich dabei entweder um ein sehr ausgefallenes Wort (z.B. Eigennamen oder ähnliches), das nicht in der Worttabelle enthalten ist, oder das Wort wurde falsch geschrieben. Wir sprechen von einem **unbekannten Wort**. Auch in diesen Fällen wird durch SITEX nicht getrennt, die Silbentrennung bleibt sicher. ISITEX erlaubt aber dem Benutzer, je nachdem welcher Fall vorliegt, das unbekannte Wort zu korrigieren oder in die Worttabelle einzutragen.

SiSiSi unterstützt eine sinnentsprechende Trennung, indem es die **Haupttrennstellen** an den Nahtstellen zusammengesetzter Wörter gegenüber den Nebentrennstellen in den Einzelwörtern bevorzugt.

Diese Bevorzugung wird durch Strafpunkte (penalties) gesteuert. Defaultmäßig erhalten die Nebentrennstellen einen höheren Wert als die Haupttrennstellen. Diese Relation kann aber von einem erfahrenen Benutzer durch den neuen TEX–Befehl `\nebenhyphenpenalty` für seine Installation geändert werden. Aber das wird für normales Arbeiten kaum notwendig sein.

3 Kurzbeschreibung des Trennverfahrens

Jedes Wort ist eine Folge von einem oder mehreren Einzelwörtern. Jedes Einzelwort besteht aus beliebig vielen Vorsilben (eventuell auch keinen), gefolgt von genau einem Stamm, abgeschlossen durch beliebig viele Endungen (ersatzweise Fugenzeichen). Eine Worttabelle enthält alle diese erwähnten Wortbestandteile (Morpheme).

Der Algorithmus sucht für jedes Wort, das eventuell getrennt werden soll, systematisch nach allen Zerlegungen, die nach der angegebenen Grammatik möglich sind. Dabei erkennt er durch Nachschauen in der Worttabelle, ob ein betrachtetes Teilstück des Wortes ein Wortbestandteil im erwähnten Sinn ist und gegebenenfalls von welcher Art es ist.

Auf diese Weise findet der Algorithmus für jede Zerlegung alle Haupttrennstellen, nämlich zwischen der letzten Endung eines Teilwortes und der folgenden Vorsilbe, bzw. dem folgenden Stamm. Außerdem erkennt er auch sofort alle Nebentrennstellen hinter den Vorsilben. Für den Rest jedes Einzelworts, bestehend aus Stamm und Endungen, muß man die Duden–Regeln für die Silbentrennung anwenden. Das sind solche Regeln wie „In einer Folge von Konsonanten ist vor dem letzten zu trennen“, z.B. tre-ten, Fül-lungen, kämp-fen. Diese Regeln sind mit all ihren Ausnahmen, z.B. Sonderbehandlung von st und ck, vollständig in *SiSiSi* eingearbeitet. Ebenso wird schon bei der Zerlegung in Einzelwörter die 3–Konsonanten–Regel berücksichtigt.

Es gibt Wörter, Fremdwörter, die nicht nach den Duden–Regeln getrennt werden, z.B. Pro-gramm oder Pan-orama. In diesen Fällen ist in der Worttabelle der Stamm als „Ausnahme“ deklariert und die möglichen Trennstellen sind angegeben. Der Algorithmus berücksichtigt natürlich solche Angaben. Eine ausführliche Beschreibung des Verfahrens findet sich in [5].

4 Mehrdeutigkeiten

Nun gibt es im Deutschen das Problem, daß viele, meist zusammengesetzte, Wörter existieren, die verschieden interpretiert und daher auch auf verschiedene Arten getrennt werden können. Das ergibt dann, sofern der Algorithmus eine unpassende Trennstelle wählt, sinnentstellende Abtrennungen, die das flüssige Lesen erschweren. Einige Beispiele:

Wachs=tube/Wach=stube

Bau=mast/Baum=ast

Personal=mangel/Person=alm=angel

Spiel=ende/spie-lende

Alle anderen Trennverfahren, die die Einzelbestandteile von zusammengesetzten Wörtern (die „Morpheme“) nicht erkennen (also auch die „pattern–Methode“), finden diese Mehrdeutigkeiten nicht. *SiSiSi* (in SITEX) hingegen nutzt solche unsicheren Trennstellen generell nicht aus. Im Beispiel „Bettruhe“ wird die angegebene Haupttrennstelle ignoriert. Nur die sichere Nebentrennstelle „Bettru-he“ kommt in Frage. Hat man aber nur wenige sichere Trennstellen in langen, zusammengesetzten Wörtern, so können in manchen Zeilen (bei Randausgleich) große Lücken bzw. „Stauchungen“ entstehen. Das Schriftbild wird unschön, häufig tritt die TEX–Warnung „underfull hbox“ auf.

Dieses Problem wird in ISITEX durch interaktives Eingreifen in den Produktionsvorgang gelöst. Trifft ISITEX auf ein mehrdeutiges Wort, so werden dem Benutzer alle möglichen Zerlegungen am Bildschirm angezeigt, sodaß er die in seinem Kontext sinngemäße auswählen kann. Diese Wahl muß für ein solches mehrdeutiges Wort nur einmal getroffen werden, d.h. bei weiteren Vorkommen desselben Wortes in demselben Dokument soll dann immer automatisch diese richtige Trennung gewählt werden. Dieser Forderung liegt die Erfahrung zugrunde, daß ein mehrdeutiges Wort praktisch nie in demselben Dokument in verschiedenen Bedeutungen vorkommt. Außerdem sind viele der von *SiSiSi* „erkannten“ Mehrdeutigkeiten im üblichen Wortschatz nicht gebräuchlich (siehe das Beispiel: Person=alm=angel). Weiters werden auch für weitere Durchläufe (nach erneutem Editieren) die richtigen Trennstellen gemerkt und wieder verwendet.

Für den Sonderfall, daß ein mehrdeutiges Wort in mehreren Bedeutungen vorkommt, z.B. „In der Wachstube liegt eine Wachstube“, gibt es die beiden folgenden Möglichkeiten:

- Man macht überhaupt nichts. Dann werden nur die sicheren Trennstellen ausgenutzt. Es entstehen keine Trennfehler, allenfalls wird das Schriftbild etwas unruhig.
- Mit dem TEX–Befehl `\-` trägt man beim Editieren bei allen (oder auch nur bei einigen ausgewählten) Vorkommen des Wortes alle Trennstellen (oder nur einen Teil davon) ein. ISITEX nutzt dann diese vorgegebenen Trennstellen aus, sucht aber in jedem Wort, das eine solche Vorgabe enthält, nicht nach weiteren Trennstellen.

Aber noch ein Vorfall muß beachtet werden. Wie kann der Benutzer, der irrtümlich für ein mehrdeutiges Wort eine falsche Trennvariante ausgewählt hat, diesen Fehler wieder beseitigen? Bei jedem weiteren Durchlauf läuft das Programm ohne Halt über alle Vorkommen dieses Wortes hinweg, es hält ja das Problem dieser Mehrdeutigkeit für gelöst. Deshalb wurde in ISITEX am Anfang der Behandlung des Dokuments ein Halt eingefügt, bei dem der Benutzer aufgefordert wird, etwaige Fehler aus vorherigen Durchläufen wieder zu beseitigen.

5 Unbekannte Wörter

Alle Wörter, für die *SiSiSi* keine Zerlegung findet, sind **unbekannte Wörter**. Das sind Wörter, die einen Schreibfehler enthalten, aber auch solche, die richtig geschrieben, dem *SiSiSi*–Algorithmus aber noch unbekannt sind (d.h. die noch nicht in der Worttabelle sind). Die einfache Lösung von SITEX lautet: Solche Worte werden nicht getrennt. Die Folgen sind die gleichen wie bei mehrdeutigen Wörtern: Lücken, Stauchungen, „underfull hbox" — und wieder aus einem für den Autor des Dokuments unerklärlichen Grund. Aber dennoch können keine Trennfehler entstehen!

Das interaktive ISITEX bleibt bei jedem unbekannten Wort stehen. Der Benutzer kann jetzt entscheiden, ob es sich um einen Tippfehler oder wirklich um ein unbekanntes Wort, d.h. um einen Namen, eine Abkürzung, ein ausgefallenes Fremdwort oder etwas Ähnliches handelt. Diese „neuen Wörter" sollen nur einmal als unbekannt gemeldet werden, also in den folgenden Produktionsläufen desselben Dokuments als bekannt akzeptiert werden. Tippfehler sollen hingegen immer angezeigt werden.

Da aber ein Eingriff in das gerade bearbeitete TEX–File nicht möglich ist (d.h. ein sofortiges Ausbessern der Tippfehler im Dokument ginge weit über den Einflußbereich der Silbentrennung hinaus und würde weitreichende Eingriffe in TEX erfordern), speichert ISITEX die vom Benutzer als „Tippfehler" bezeichneten Wörter in einem File, das man beim nächsten Editieren als Gedächtnisstütze verwenden kann.

Welche neuen Möglichkeiten ISITEX bietet, ist zusammenfassend in Abbildung 1 dargestellt.

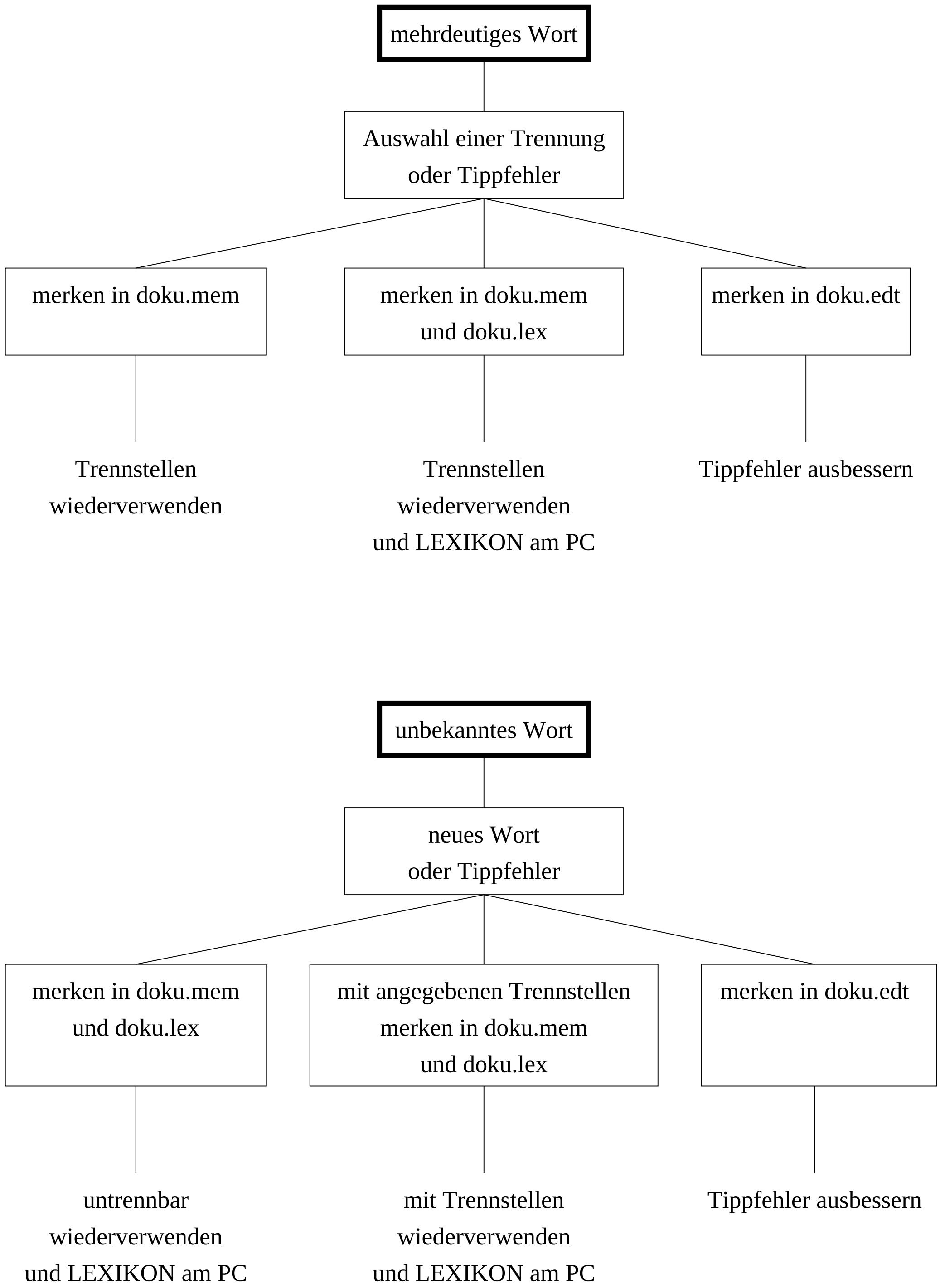

Abbildung 1: Die neuen Möglichkeiten

6 ISITEX intern: Die neuen Files

Um die obigen Erweiterungen zu ermöglichen, werden die entsprechenden Informationen in drei Files gespeichert. „doku“ steht im Folgenden für den Namen des gerade bearbeiteten Dokuments.

- **doku.mem**: Der Benutzer wählt bei mehrdeutigen Wörtern die passende Bedeutung aus. Diese Wahl wird in `doku.mem` gespeichert. Von jetzt an werden alle Vorkommnisse dieses Wortes entsprechend dieser Festlegung getrennt. Achtung! Kommt ein Wort mit verschiedenen Endungen im Text vor, so wird für jede Variante extra gefragt. (Kommt also in einem Dokument zB. dreimal „Wachstube“ und fünfmal „Wachstuben“ vor, so muß der Benutzer je einmal die richtige Trennung für „Wachstube“ und „Wachstuben“ angeben.) Ebenso werden alle Wörter, die dem System unbekannt sind und für das Dokument spezielle Fremdwörter oder Namen darstellen, durch ein Kommando des Benutzers (eventuell mit den eingegebenen Trennstellen) in `doku.mem` gespeichert. Diese Information wird dann bei allen weiteren ISITEX Produktionsläufen berücksichtigt und entsprechend erweitert (wenn beim weiteren Editieren neue mehrdeutige oder unbekannte Wörter hinzugefügt wurden).

- **doku.lex**: Bei mehrdeutigen und insbesondere bei unbekannten Wörtern, kann sich der Benutzer dazu entschließen, daß sie (zusätzlich zu der erwähnten Eintragung in `doku.mem`) in die Worttabelle von *SiSiSi* eingetragen werden sollen. Unbekannte Wörter kommen daher automatisch und mehrdeutige Wörter auf Wunsch in das File `doku.lex`, damit sie leicht mit dem „*LEXIKON*–System“ am PC in die Worttabelle eingetragen werden können. So ist es dem Benutzer möglich, eine spezielle Worttabelle für seinen persönlichen Wortschatz zu schaffen.

- **doku.edt**: Alle als Tippfehler erkannten Wörter werden in das File `doku.edt` geschrieben. Dieses File unterscheidet sich grundsätzlich von den beiden obigen. Eintragungen in diesem File werden nicht automatisch für weitere Vorkommen desselben Wortes und auch nicht für weitere Durchläufe verwendet. Es ist als „Merkfile“ des Benutzers für das nächste Editieren gedacht. Er kann es daher anschauen, ausdrucken, editieren und löschen. Weiters wird dieses File bei jedem Durchlauf neu angelegt. Das ist sinnvoll, weil Tippfehler ja nicht hingenommen werden sollen (weder von ISITEX noch vom Benutzer). Es werden also immer alle Tippfehler gemeldet. Das wiederholte Aufmerksammachen ist durchaus beabsichtigt.

Die Files `doku.mem` und `doku.lex` dürfen nicht vom Benutzer verändert werden, weil sie vom System wieder verwendet werden. `doku.edt` darf geändert oder zerstört werden.

7 Wie erzeugt man ein druckfertiges Manuskript mit ISITEX?

Beim ersten Editieren bereitet man das Manuskript genauso wie für „normales“ TEX auf. Die Befehle für die Fonts, die Formate, usw. fügt man in den Text ein.

Beim ersten Durchlauf durch das ISITEX–System beantwortet man die Fragen des Systems bezüglich mehrdeutiger und unbekannter Wörter sinnvoll. Es entstehen die Files `doku.mem`, `doku.lex` und `doku.edt` (siehe Abbildung 2).

Im allgemeinen wird man auch das entstandene File drucken, um weitere Hinweise für Korrekturen im Manuskript zu erhalten. In diesem Ausdruck sind schon alle Trennstellen berücksichtigt, die man — wie in Kapitel 6 beschrieben — an das File `doku.mem` gemeldet hat.

Jetzt kann man die Worttabelle mit dem System *LEXIKON* erweitern. Man wird (siehe Kapitel 10) alle speziellen Wörter dieses Dokuments, bzw. des Fachvokabulars eintragen. *LEXIKON* benutzt dabei das File `doku.lex`.

Danach erfolgt das nächste Editieren. Mit den Hinweisen aus dem File `doku.edt` und den Erkenntnissen vom Korrekturlesen wird man insbesondere alle Tippfehler korrigieren, aber auch andere Korrekturen im Manuskript vornehmen.

Beim folgenden nächsten Durchlauf bleibt das System nicht mehr bei allen mehrdeutigen und unbekannten Wörtern, die in einem früheren Durchlauf in `doku.mem` eingetragen wurden, stehen. Am Anfang dieses Durchlaufs gibt es jedoch Gelegenheit, frühere falsche Festlegungen zu löschen. Neue mehrdeutige oder unbekannte Wörter (auch alte Tippfehler) werden wie im ersten Durchgang behandelt.

Diese Vorgänge werden solange wiederholt, bis die Qualität des gedruckten Manuskripts den Ansprüchen genügt.

Die Wege, die die Daten nehmen, sind in Abbildung 2 zusammenfassend dargestellt.

8 „xxx“ ist ein mehrdeutiges Wort! — was tun?

Wenn ISITeX ein mehrdeutiges Wort findet, dann sieht das folgendermaßen aus (am Beispiel des Wortes „Wachstube“):

```
'wachstube' ist ein mehrdeutiges Wort!
 0: nur die sicheren Trennstellen waehlen ('wachstu-be')
 1: 'wach=stu-be'
 2: 'wachs=tu-be'
 9: Tippfehler! - merken im File 'doku.edt'
Bitte waehlen Sie eine Bedeutung z.B. '0'+'Enter' aus:
 (zusaetzlich E z.B. '0E'+'Enter':
  Wort auch fuer LEXIKON vormerken)
```

Nun muß man eine Ziffer gefolgt von „Enter“ eingeben. Ist z.B. die Bedeutung „Wach=stube“ die gemeinte, so gibt man „1“+„Enter“ ein. Ab jetzt wird das Wort „Wachstube“ im ganzen Dokument und auch in künftigen Durchläufen entsprechend getrennt. Tritt der überaus seltene Fall auf, daß ein Wort tatsächlich in mehreren Bedeutungen in ein und demselben Dokument vorkommt, so kann man sich durch „0“+„Enter“ auf die sicheren Trennstellen beschränken. Will man das Wort später auch in die Worttabelle eintragen, so gibt man zusätzlich noch ein „E“ an (also zB. „1E“+„Enter“ statt „1“+„Enter“). Dadurch wird das Wort zusätzlich in das File `doku.lex` eingetragen, das die Arbeit mit dem Lexikon–System erleichtert. Gibt man „9“+„Enter“ ein, so wird das Wort nur (!) ins File `doku.edt` geschrieben und kann dann beim Editieren von `doku.tex` gesucht und ausgebessert werden.

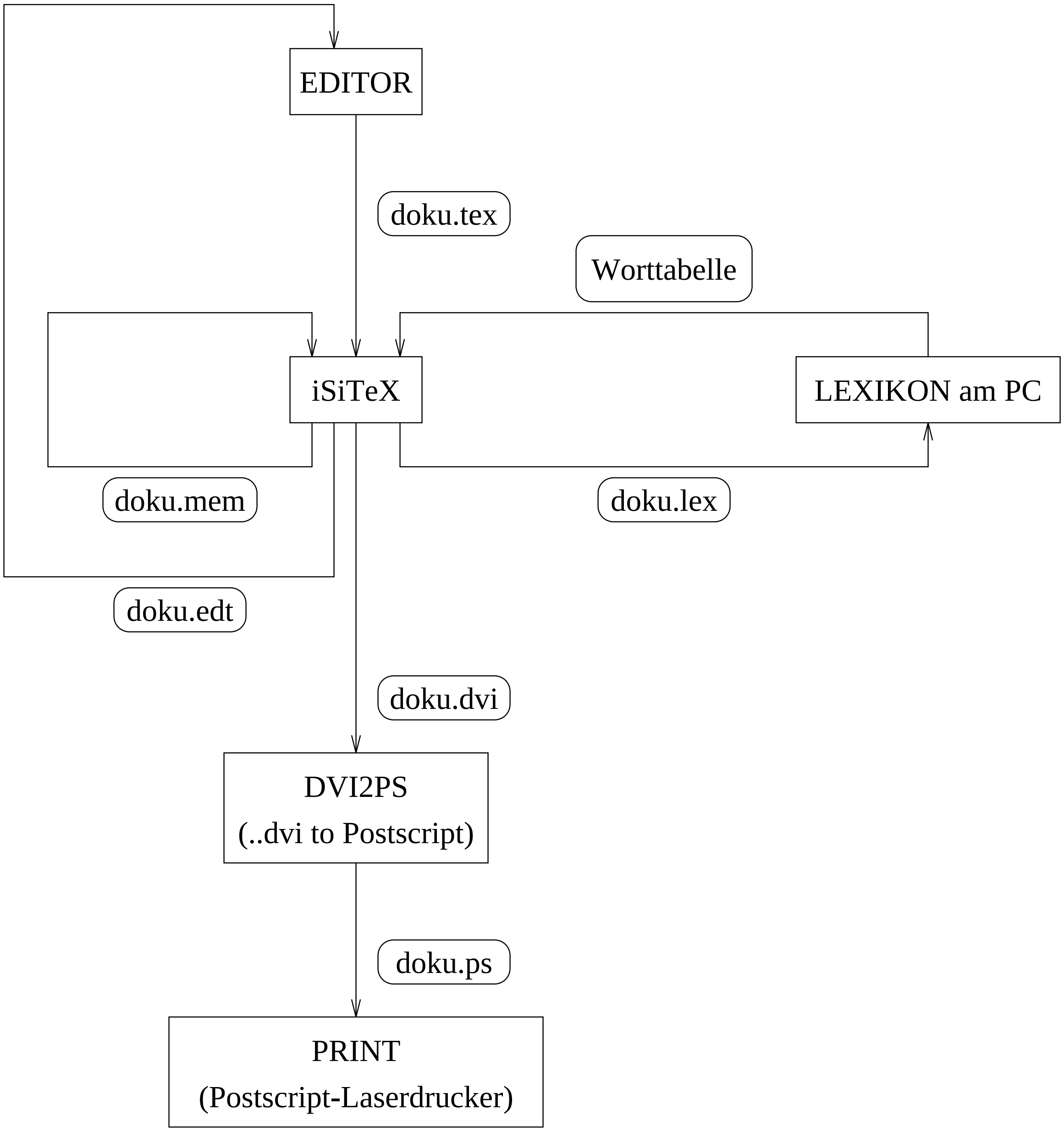

Abbildung 2: Die Wege der Daten

9 „xxx“ ist ein unbekanntes Wort! — was tun?

Wenn ISITEX ein unbekanntes Wort findet, dann sieht das folgendermaßen aus (am Beispiel des Wortes „Acetylsalicylsäure“):

```
'acetylsalicyls"aure' ist ein unbekanntes Wort!
 0: neues Wort, merken in 'doku.lex'
 1: neues Wort, Trennstellen angeben und merken in 'doku.lex'
 9: Tippfehler! - merken im File 'doku.edt'
Bitte waehlen Sie eine Bedeutung (0,1,9) aus:
```

Nun muß man eine Ziffer gefolgt von „Enter“ eingeben. Ist das Wort beispielsweise ein kurzes Fremdwort, so gibt man „0“+„Enter“ ein. Ab jetzt ist das Wort bekannt; es wird jedoch nicht getrennt. Es wird in den Files `doku.mem` und auch `doku.lex` gespeichert, um es nachher eventuell mit dem *LEXIKON*–System leicht in die Worttabelle eintragen zu können. Handelt es sich bei dem unbekannten Wort aber um ein langes (zusammengesetztes) fremdes Wort, so ist es — aus den oben genannten Gründen — sinnvoll, Trennstellen vorzugeben. Zu diesem Zweck tippt man „1“+„Enter“. Daraufhin erscheint dann folgende Eingabeaufforderung:

```
Bitte den Trennvektor angeben
 ('-' Nebentrennstelle, '=' Haupttrennstelle, '.' Weiterruecken,
 Trennstrich unter dem Buchstaben angeben, hinter dem
 getrennt werden soll):
acetylsalicyls"aure
```

Nun kann man direkt unter dem Wort die Trennstellen angeben:

```
acetylsalicyls"aure
..-..=.-.-..=...-
```

Nachdem man „Enter“ gedrückt hat, ist „acetylsalicylsäure“ bekannt und wird gegebenenfalls „ace-tyl=sa-li-cyl=säu-re“ getrennt.

Als dritte Wahlmöglichkeit hat man noch „9“+„Enter“. Dann wird das Wort ins File „doku.edt“ geschrieben und kann dann beim Editieren von „doku.tex“ gesucht und ausgebessert werden.

10 Änderung der Worttabelle

Es ist zu empfehlen, spezielle Wörter des benutzereigenen Vokabulars auch wirklich in die Worttabelle einzutragen. *SiSiSi* kennt nämlich nach dem Eintragen eines solchen neuen Wortes auch alle Abwandlungen dieses Wortes, die durch Voransetzen von Vorsilben oder auch Anfügen von Endungen entstehen, sowie alle seine Kombinationen mit anderen Wörtern. Natürlich erkennt *SiSiSi* diese Wörter dann alle richtig und trennt sie bei Bedarf korrekt. Beispielsweise sind keine Tiernamen in der normalen Worttabelle. Trägt der Benutzer, der ein Tierbuch schreibt, Fuchs als Stamm in die Tabelle ein, dann behandelt *SiSiSi* auch Fuchses, Silberfuchs, Fuchsschwanz und „es fuchst mich" korrekt wie jedes andere Wort. Allerdings müssen Füchse wegen der Wandlung des u in ein ü gesondert eingetragen werden.

Man sollte nur neue Wortstämme eintragen. Bei Fremdwörtern muß man jedoch darauf achten, ob eine Ausnahme zu den Duden–Regeln vorliegt. Dazu muß man sich naturgemäß mit dieser Problematik befassen, z.B. [5], [6] und [9]. Ausnahmen sind z.B. Pro-gramm, Kata-strophe, Pin-up-girl, Pan-ora-ma. Man muß recht sorgfältig arbeiten.

Man kann folgende Morphem-Arten in die Worttabelle eintragen:

	Art	Kurzbezeichnung	Erklärung
1	Stamm	s	normaler Stamm, wird nach Duden–Regeln getrennt
2	Stamm, nicht geeignet für Wortzusammensetzungen	st	Wie 1, jedoch kann dieser Stamm nicht mit anderen zu einem zusammengesetzten Wort verknüpft werden
3	Ausnahme	a i j oder a i	Es wird abweichend von den Dudenregeln nach dem i-ten und j-ten Buchstaben getrennt (und nur dort)
4	Ausnahme mit Ende–Markierung	au i k oder au k	Wie 3, jedoch definiert k keine Trennstelle, nach dem k-ten Buchstaben setzen die Duden-Regeln wieder ein

	Art	Kurzbezeichnung	Erklärung
5	Vorsilbe	v	Nach der Vorsilbe wird getrennt
6	Endung	e	Die Endung(en) wird mit dem Stamm zusammen den Duden–Regeln unterworfen

Die Änderung der Worttabelle kann man auf eine der folgenden Arten durchführen:

a) mit dem mitgelieferten Lexikon-System auf einem PC
Auf dem PC unterstützt Sie ein Dialogsystem beim Einfügen der neuen Wörter. Sie brauchen sich um Einzelheiten der Kodierung nicht zu kümmern. Außerdem zeigt Ihnen das System in jedem Fall zur Kontrolle an, wie *SiSiSi* von nun an die fraglichen Wörter trennen wird.

Nach der Änderung müssen Sie die Worttabelle worttab.htx mit dem htx2tex–Programm in die unter UNIX übliche Kodierung (4 für ä, ..., 1 für ß) umwandeln. Sie erhalten dadurch das neue File sihyphen.tex.

b) durch Editieren des Files sihyphen.tex
Sie können unmittelbar Eintragungen in der Art vornehmen, wie sie in obiger Tabelle beschrieben sind. Sie müssen dann allerdings die (benutzerfreundliche ?) Kodierung von UNIX beachten und erhalten auch keine explizite Angabe darüber, wie *SiSiSi* die neuen Wörter später trennen wird.

Nach Abschluß dieser Änderungsarbeiten ist in beiden Fällen a) und b) SITEX bzw. ISITEX mit initex neu zu initialisieren, so wie das in der Installationsanleitung von SITEX bzw. ISITEX für UNIX beschrieben ist.

Literatur

[1] Knuth, Donald E.: TEX: The Program, Addison Wesley, Computers and Typesetting, Vol. B

[2] Knuth, Donald E.: The TEXbook, Addison Wesley

[3] Lamport, Leslie: LATEX. A document preparation system, Addison Wesley

[4] Kopka, Helmut: LATEX: eine Einführung, Addison Wesley

[5] Barth W., Nirschl H.: Implementierung eines Verfahrens für die Silbentrennung, TU Wien, Institut für Praktische Informatik, Bericht Nr. 26, 1985

[6] Barth W., Nirschl H.: Sichere sinnentsprechende Silbentrennung für die deutsche Sprache. Angewandte Informatik 4, S. 152–159 (1985)

[7] Barth W., Steiner H.: Deutsche Silbentrennung für TEX 3.1, DANTE, Die TEXnische Komödie, Heft 1/1992

[8] *SiSiSi* für TEX, Installationsanleitung, über anonymes FTP am server `eiunix.tuwien.ac.at`.

[9] DUDEN, Rechtschreibung der deutschen Sprache und der Fremdwörter. 19. Auflage, Bibliographisches Institut, Mannheim/Wien/Zürich (1986)