BEX – Binary Encoded XML (Document)

Binary Encoded XML (Document) beschreibt das binäre Datenformat zur redundanzarmen Abbildung eines XML Dokuments. Ziel dieses Formats ist es, eine leichtgewichtige, nur lesende DOM-Implementation darauf aufsetzen zu können, welche signifikant weniger Arbeitsspeicher verbraucht, als eine zumeist auch schreiben könnende Implementation einer Standard XML Softwarebibliothek. Diese leichtgewichtige DOM-Implementation kann die Binärdatei Blockweise in den Arbeitsspeicher laden und dort zur Wiederverwendung gemäß einer most-recently-used Strategie vorhalten. Die offset-Listen sollten dazu vollständig im Arbeitsspeicher vorgehalten werden.

# Datenstruktur: BEX

Kodiert einen DOM Dokumentknoten mit seinen enthaltenen Text-, Element- und Attributknoten als binäre Datenstruktur zum wahlfreien Zugriff.

| Feld | Format | Anzahl | Beschreibung |
| --- | --- | --- | --- |
| attrUriPool | TextValuePool | 1 | Auflistung der Zeichenketten für die URIs der Attributknoten. |
| attrNamePool | TextValuePool | 1 | Auflistung der Zeichenketten für die Namen der Attributknoten. |
| attrValuePool | TextValuePool | 1 | Auflistung der Zeichenketten für die Werte der Attributknoten. |
| attrGroupPool | AttrGroupPool | 1 | Auflistung der Attributknotenlisten. |
| elemUriPool | TextValuePool | 1 | Auflistung der Zeichenketten für die URIs der Elementknoten. |
| elemNamePool | TextValuePool | 1 | Auflistung der Zeichenketten für die Namen der Elementknoten. |
| elemValuePool | TextValuePool | 1 | Auflistung der Zeichenketten für die Werte in Elementknoten. |
| elemGroupPool | ElemGroupPool | 1 | Auflistung der Kindknotenlisten. |
| documentGroupRef | INT1..4 | 1 | Kindknotenliste des Dokumentknoten. Ist (i+1+elemValuePool.size) als Referenz auf die i-te Kindknotenliste in elemGroupPool. |

# Datenstruktur: TextValuePool

Kodiert eine Auflistung von Zeichenketten.

| Feld | Format | Anzahl | Beschreibung |
| --- | --- | --- | --- |
| size | INT4 | 1 | Anzahl der Zeichenketten. |
| length | INT1 | 1 | Anzahl der Byte je Startposition. |
| offset | INT1..4 | size | Startpositionen der Zeichenketten im Speicherbereich item. Das erste Byte der i-ten Zeichenkette liegt bei Position offset[i] und das nach dem letzten bei offset[i+1]. Die Startposition offset[0] ist implizit 0. |
| item | TextValueItem | size | Speicherbereich mit den Bytes der Zeichenketten. |

# Datenstruktur: TextValueItem

Kodiert eine Zeichenkette.

| Feld | Format | Anzahl | Beschreibung |
| --- | --- | --- | --- |
| size | - | - | Anzahl der Bytes in data. |
| data | INT1 | size | Bytes der UTF-8-kodierten Zeichenkette. |

# Datenstruktur: AttrGroupPool

Kodiert eine Auflistung von Attributknotenlisten.

| Feld | Format | Anzahl | Beschreibung |
| --- | --- | --- | --- |
| size | INT4 | 1 | Anzahl der Attributknotenlisten. |
| length | INT1 | 1 | Anzahl der Byte je Startposition. |
| offset | INT1..4 | size | Startpositionen der Attributknotenlisten im Speicherbereich item. Der erste Attributknoten der i-ten Attributknotenliste liegt bei Position offset[i] und der nach dem letzten bei offset[i+1]. Die Startposition offset[0] ist implizit 0. |
| item | AttrGroupItem | size | Speicherbereich mit den Attributknoten aller Attributknotenlisten. |

# Datenstruktur: AttrGroupItem

Kodiert eine Attributknotenliste.

| Feld | Format | Anzahl | Beschreibung |
| --- | --- | --- | --- |
| size | - | - | Länge der Liste. |
| node | AttrGrpupNode | size | Attributknoten der Liste. |

# Datenstruktur: AttrGroupNode

Kodiert einen Attributknoten.

| Feld | Format | Anzahl | Beschreibung |
| --- | --- | --- | --- |
| uriRef | INT1..4 | 0..1 | Ist abwesend, wenn der attrUriPool leer ist. Ist (0), wenn der Attributknoten keinen URI besitzt. Ist (i+1) als Referenz auf den i-ten URI im attrUriPool. |
| nameRef | INT1..4 | 1 | Ist (i) als Referenz auf den i-ten Namen im attrNamePool. |
| valueRef | INT1..4 | 1 | Ist (i) als Referenz auf den i-ten Wert im attrValuePool. |

# Datenstruktur: ElemGroupPool

Kodiert eine Auflistung von Kindknotenlisten.

| Feld | Format | Anzahl | Beschreibung |
| --- | --- | --- | --- |
| size | INT4 | 1 | Anzahl der Kindknotenlisten. |
| length | INT1 | 1 | Anzahl der Byte je Startposition. |
| offset | INT1..4 | size | Startpositionen der Kindknotenlisten im Speicherbereich item. Der erste Kindknoten der i-ten Kindknotenliste liegt bei Position offset[i] und der nach dem letzten bei offset[i+1]. Die Startposition offset[0] ist implizit 0. |
| item | ElemGroupItem | size | Speicherbereich mit den Kindknoten der Kindknotenlisten. |

# Datenstruktur: ElemGroupItem

Kodiert eine Auflistung von Kindknotenlisten.

| Feld | Format | Anzahl | Beschreibung |
| --- | --- | --- | --- |
| size | - | - | Länge der Liste. |
| node | ElemGrpupNode | size | Kindknoten der Liste. |

# Datenstruktur: ElemGroupNode

Kodiert einen Kindknoten meist als Element- und sehr selten als Textknoten.

| Feld | Format | Anzahl | Beschreibung |
| --- | --- | --- | --- |
| uriRef | INT1..4 | 0..1 | Ist abwesend, wenn der elemUriPool leer ist. Ist (0), wenn der Kindknoten ein Textknoten oder ein Elementknoten ohne URI ist. Ist (i+1) als Referenz auf den i-ten URI im elemUriPool, wenn der Kindknoten ein Elementknoten ist. |
| nameRef | INT1..4 | 1 | Ist (0), wenn der Kindknoten ein Textknoten ist. Ist (i+1) als Referenz auf den i-ten Namen im elemNamePool, wenn der Kindknoten ein Elementknoten ist. |
| contentRef | INT1..4 | 1 | Ist (0), wenn der Kindknoten ein Elementknoten ohne Kindknoten ist. Ist (i+1) als Referenz auf den i-ten Wert im elemValuePool, wenn der Kindknoten ein Textknoten oder ein kindelementloser Elementknoten ist. Ist (i+1+elemValuePool.size) als Referenz auf die i-te Kindknotenliste im elemGroupPool, wenn der Kindknoten ein Elementknoten mit Kindelementen ist. |
| attributesRef | INT1..4 | 0..1 | Ist abwesend, wenn der attrGroupPool leer ist. Ist (0), wenn der Kindknoten ein Textknoten oder ein Elementknoten ohne Attributknoten ist. Ist (i+1) als Referenz auf die i-te Attributknotenliste im attrGroupPool, wenn der Kindknoten ein Elementknoten mit Attributknoten ist. |