

PDF Parsing 가이드

- 실행 순서
- 검수 시 참고사항

parser_samsung/	# 최상위 폴더
├── Readme.md	# 사용 방법
├── labeling_guide.pptx	# 라벨링 가이드
├── setup_parsing_samsung.txt	# 파이썬 및 라이브러리 버전
├── main.py	# 메인 코드 (pdf parser 생성시 실행)
├── utils/	# 유틸 파일 폴더
├── utils_pdf.py	# pdf 로드 및 pdf 데이터 추출 코드
├── utils_yolo.py	# pdf 레이아웃 탐지용 코드 (라벨링이 되어 있으면 생략)
└── utils_parsing.py	# 객체별 순서 및 parsing data 생성시 필요한 각종 후처리 알고리즘 코드
├── weights/	# yolo 모델 웨이트 폴더
└── best.pt	# (현재 yolov11x 모델 사용)
├── move_data.py	# pdf 레이아웃 데이터 라벨링 코드
├── pdf_to_image.py	# 새로운 pdf를 이미지로 변화하는 코드
├── datasets_labels.py	# 새로운 pdf를 이미지로 변환 한 다음(수동으로 진행 필요), 자동으로 라벨링하는 코드
└── line_merge.py	# parsing data 검수 전, 줄바꿈 부분을 띄어쓰기로 처리하는 코드

PDF Parsing 가이드 – 실행 순서

실행 순서

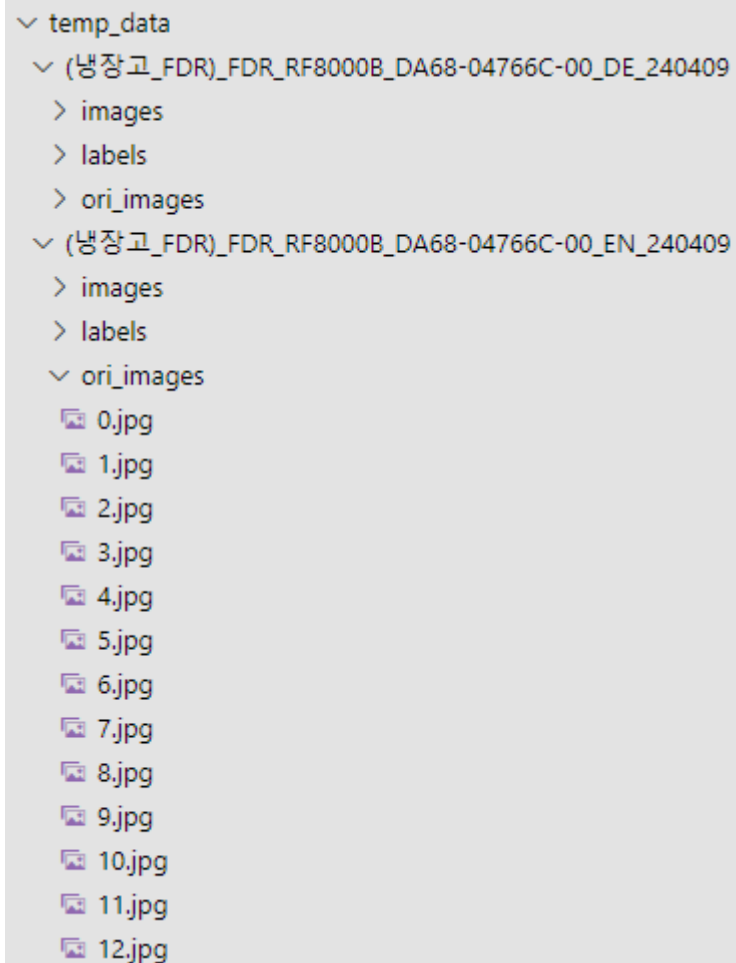
1. PDF 레이아웃 라벨링
 - 1) PDF 파일 업로드
 - 2) 'pdf_to_image.py' 코드 실행
 - 3) 'datasets_labels.py' 코드 실행
2. 라벨링 툴을 이용하여 라벨링 결과 검수
 - 1) 'move_data.py' 코드 실행
3. PDF Parsing Data 생성
 - 1) 'main.py' 코드 실행
4. Parsing Data 검수 전 전처리
 - 1) 'line_merge.py' 코드 실행
5. Parsing Data 검수
 - 1) parsing 결과 검수

parser_samsung/	# 최상위 폴더
├── Readme.md	# 사용 방법
├── labeling_guide.pptx	# 라벨링 가이드
├── setup_parsing_samsung.txt	# 파이썬 및 라이브러리 버전
├── main.py	# 메인 코드 (pdf parser 생성시 실행)
├── utils/	# 유틸 파일 폴더
├── utils_pdf.py	# pdf 로드 및 pdf 데이터 추출 코드
├── utils_yolo.py	# pdf 레이아웃 탐지용 코드 (라벨링이 되어 있으면 생략)
└── utils_parsing.py	# 객체별 순서 및 parsing data 생성시 필요한 각종 후처리 알고리즘 코드
├── weights/	# yolo 모델 웨이트 폴더
└── best.pt	# (현재 yolov11x 모델 사용)
├── move_data.py	# pdf 레이아웃 데이터 라벨링 코드
├── pdf_to_image.py	# 새로운 pdf를 이미지로 변환하는 코드
├── datasets_labels.py	# 새로운 pdf를 이미지로 변환 한 다음(수동으로 진행 필요), 자동으로 라벨링하는 코드
└── line_merge.py	# parsing data 검수 전, 줄바꿈 부분을 띄어쓰기로 처리하는 코드

PDF Parsing 가이드 – 실행 순서

1. PDF 레이아웃 라벨링

- 1) Parsing 하고자 하는 삼성전자 가전제품 설명서 PDF를 'pdf/' 폴더에 저장
- 2) 'pdf_to_image.py' 코드를 실행하여 'temp_data/{pdf이름}/ori_images/' 폴더에 PDF를 시작페이지 0번 부터 '0.jpg'로 순차적으로 저장
- 3) 'datasets_labels.py' 코드를 실행하여 'YOLOv12x' 모델로 각 페이지별 레이아웃 탐지를 진행
 - '--label_path'와 '--pdf_path'에 라벨링 폴더 경로 및 PDF 폴더 경로 지정
 - ex) --label_path temp_data --pdf_path pdf
 - 라벨값을 'temp_data/{pdf이름}/labels/'에 '{페이지번호}.txt' 로 저장
 - 결과 이미지를 'temp_data/{pdf이름}/images/'에 '{페이지번호}.jpg' 로 저장
 - 코드 실행 시 기존 결과를 덮어쓰기에 주의해야 함



PDF Parsing 가이드 – 실행 순서

2. 라벨링 툴을 이용하여 라벨링 결과 검수

1) 'move_data.py' 코드 실행

- 라벨링 툴 가이드 'labeling_guide.pdf' 를 참고



PDF Parsing 가이드 – 실행 순서

3. PDF Parsing Data 생성

1) 'main.py' 코드 실행

- 'pdf_to_image' 폴더에 PDF 파일명 폴더에 PDF 페이지 저장
- 'pdf_results' 폴더에 parsing 결과를 생성하기 위한 데이터들이 저장
- 'pdf_parser_results' 폴더에 Pdf Parsing Data 파일 생성

```
parser_samsung/           # 최상위 폴더
├─ pdf_parser_results/    # parsing data 생성 폴더
│   ├── pdf1.xlsx         # pdf1에 대한 최종 parsing data 파일
│   ├── pdf2.xlsx         #
│   └─ ...
└─ pdfn.xlsx              #
├─ pdf_results/           # pdf 추출 데이터
│   ├── debug_regions/   # 영역 확인용 디버그 폴더
│   ├── pdf_elements/    # 페이지별 이미지, 표, 아이콘 이미지 모음 폴더
│   ├── pdf_refined_text/ # 페이지별 최종 수정된 텍스트 및 영역 순서
│   ├── pdf_results_image/ # 페이지별 영역 탐지 결과
│   ├── pdf_text/        # 페이지별 pdf 로더를 통한 텍스트 추출
│   └─ pdf_yaml/         # 페이지별 최종 수정 전 텍스트 및 영역 순서
└─ pdf_to_image/         # pdf 페이지를 이미지로 변환
```

```
▼ pdf_parser_results
  (냉장고_FDR)_FDR_RF8000B_DA68-04766C-00_DE_240409_dqa.xlsx
  (냉장고_FDR)_FDR_RF8000B_DA68-04766C-00_EN_240409_dqa.xlsx
▼ pdf_results
  > debug_regions
  > pdf_elements
  > pdf_refined_text
  > pdf_results_image
  > pdf_text
  > pdf_yaml
▼ pdf_to_image
  > (냉장고_FDR)_FDR_RF8000B_DA68-04766C-00_DE_240409
  > (냉장고_FDR)_FDR_RF8000B_DA68-04766C-00_EN_240409
```

PDF Parsing 가이드 – 실행 순서

4. Parsing Data 검수 전 전처리

1) 'line_merge.py' 코드를 실행하여 'pdf_parser_results' 폴더에 있는 PDF Parsing Data 의 모든 행의 줄바꿈을 규칙에 따라 일괄적으로 제거

SCHRITT 2 Bodenbelag	die gewünschte Stelle.	참고	41
SCHRITT 3 Ausbauen der Tür beim Transport	Wenn die Kühl-/Gefrierkombination aufgrund seiner Größe nicht durch den Eingang Ihres Hauses oder Ihrer Küche passt, befolgen Sie diese Anweisungen, um die Türen der Kühl-/Gefrierkombination zuerst aus- und dann wieder einzubauen. Wenn Sie die Türen nicht ausbauen müssen, setzen Sie den Installationsvorgang ab Seite 48 fort.	일반	42
SCHRITT 3 Ausbauen der Tür beim Transport	Ziehen Sie das Netzkabel der Kühl-/Gefrierkombination aus der Steckdose, bevor Sie die Türen ausbauen.	경고	42
Ausbauen der Türen des Kühlabteils	1. Entfernen Sie bei geschlossenen Türen die (3) Schrauben an der oberen Abdeckung mit einem Kreuzschlitzschraubendreher. Ziehen Sie die obere Abdeckung mit den Anschlüssen nach oben heraus.	일반	42
Ausbauen der Türen des Kühlabteils	2. Trennen Sie die beiden Kabelanschlüsse an der linken Tür. Die Anordnung der Anschlüsse hängt vom jeweiligen Modell ab.	일반	43
Ausbauen der Türen des Kühlabteils	• Entfernen Sie nicht beide Türen gleichzeitig. Achten Sie darauf, jeweils eine Tür nach der anderen zu entfernen. Andernfalls kann eine der beiden Türen hinfallen und Verletzungen verursachen. • Stellen Sie sicher, dass das Kühlabteil ausgeschaltet ist, bevor Sie die Stecker aus der Dose ziehen.	주의	43
Ausbauen der Türen des Kühlabteils	3. Trennen Sie den Wasserschlauch vom Anschluss an der linken Tür (siehe Abbildung). - Dieser Schritt gilt nicht für die Modelle ohne AutoFill-Krug und Wasserspender.	주의	43
Ausbauen der Türen des Kühlabteils	4. Ziehen Sie die Klemme in mehreren Schritten vorsichtig auf. Um Verletzungen der Finger zu vermeiden, dürfen Sie nicht plötzlich eine Kraft auf die Klemme ausüben.	주의	43

줄바꿈 전

SCHRITT 2 Bodenbelag	Pappe oder einen alten Teppich, und ziehen Sie es damit an die gewünschte Stelle.	참고	41
SCHRITT 3 Ausbauen der Tür beim Transport	Wenn die Kühl-/Gefrierkombination aufgrund seiner Größe nicht durch den Eingang Ihres Hauses oder Ihrer Küche passt, befolgen Sie diese Anweisungen, um die Türen der Kühl-/Gefrierkombination zuerst aus- und dann wieder einzubauen. Wenn Sie die Türen nicht ausbauen müssen, setzen Sie den Installationsvorgang ab Seite 48 fort.	일반	42
SCHRITT 3 Ausbauen der Tür beim Transport	Ziehen Sie das Netzkabel der Kühl-/Gefrierkombination aus der Steckdose, bevor Sie die Türen ausbauen.	경고	42
Ausbauen der Türen des Kühlabteils	1. Entfernen Sie bei geschlossenen Türen die (3) Schrauben an der oberen Abdeckung mit einem Kreuzschlitzschraubendreher. Ziehen Sie die obere Abdeckung mit den Anschlüssen nach oben heraus.	일반	42
Ausbauen der Türen des Kühlabteils	2. Trennen Sie die beiden Kabelanschlüsse an der linken Tür. Die Anordnung der Anschlüsse hängt vom jeweiligen Modell ab.	일반	43
Ausbauen der Türen des Kühlabteils	• Entfernen Sie nicht beide Türen gleichzeitig. Achten Sie darauf, jeweils eine Tür nach der anderen zu entfernen. Andernfalls kann eine der beiden Türen hinfallen und Verletzungen verursachen. • Stellen Sie sicher, dass das Kühlabteil ausgeschaltet ist, bevor Sie die Stecker aus der Dose ziehen.	주의	43
Ausbauen der Türen des Kühlabteils	3. Trennen Sie den Wasserschlauch vom Anschluss an der linken Tür (siehe Abbildung). - Dieser Schritt gilt nicht für die Modelle ohne AutoFill-Krug und Wasserspender.	주의	43
Ausbauen der Türen des Kühlabteils	4. Ziehen Sie die Klemme in mehreren Schritten vorsichtig auf. Um Verletzungen der Finger zu vermeiden, dürfen Sie nicht plötzlich eine Kraft auf die Klemme ausüben.	주의	43
Ausbauen der Türen des Kühlabteils	5. Ziehen Sie die Klemme heraus, um sie zu entfernen.	주의	43
Ausbauen der Türen des Kühlabteils	6. Öffnen Sie die Tür im 90 Grad-Winkel. Halten Sie die Tür fest und ziehen Sie das obere Scharnier nach oben heraus. Achten Sie beim Entfernen des Scharniers darauf, die Verbindungskabel nicht zu beschädigen.	주의	44
Ausbauen der Türen des Kühlabteils	7. Ziehen Sie die Tür vorsichtig nach oben heraus. Vermeiden Sie übermäßige Krafteinwirkung auf die Tür. Die Tür kann herunterfallen und Verletzungen verursachen.	주의	44
Ausbauen der Türen des Kühlabteils	8. Legen Sie die Tür auf eine ebene Fläche.	주의	44
Ausbauen der Türen des Kühlabteils	9. Wiederholen Sie mit der anderen Tür die Schritte 1-8, mit Ausnahme von Schritt 3. An der Tür auf der rechten Seite befindet sich kein Wasserschlauch. - Schritt 2 gilt nicht für die Modelle ohne Vitrintür.	주의	44
Ausbauen der Türen des Kühlabteils	So bauen Sie die Türen des Kühlabteils wieder ein	일반	45

줄바꿈 후

PDF Parsing 가이드 – 실행 순서

5. Parsing Data 검수

- 1) 'pdf_parser_results' 폴더에 있는 PDF Parsing Data 를 해당 데이터의 삼성전자 가전제품 설명서 PDF와 대조하면서 검수 작업을 진행
- 2) 해당 제품의 'category', 'products', 'target_device(kr)', 'target_device(en)', 'model_names', 'product_names' 정보를 직접 입력

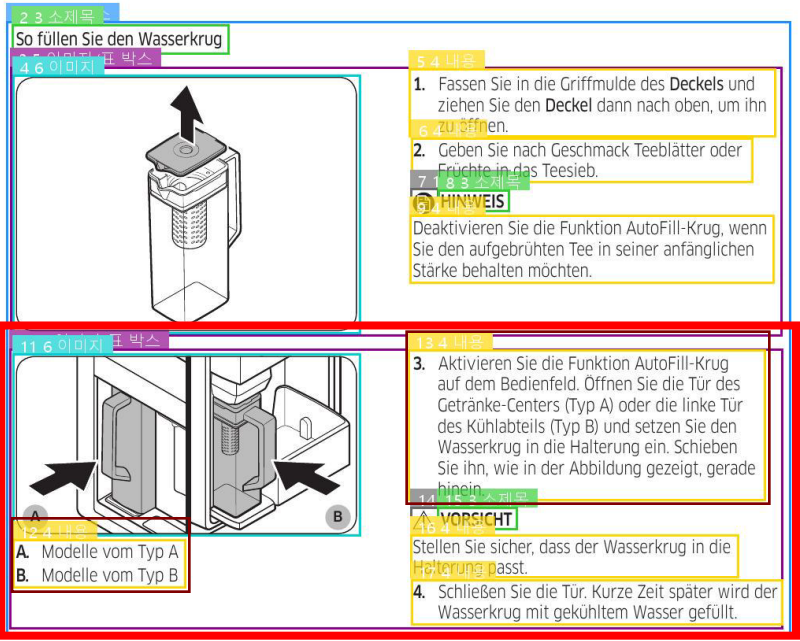
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	No	category	products	target_device(kr)	target_device(en)	model_names	product_names	title	subtitle	lasttitle	content	content_type	page
2	1							Sicherheitsinformationen			Lesen Sie dieses Handbuch sorgfältig durch, um sicherzustellen, dass Sie die Merkmale	일반	4
3	2							Sicherheitsinformation: Wichtige Informationen zu den Sicherheitshin	• Die Warnungen und wichtigen Sicherheitsinformationen		in diesem Benutzerhandbu	일반	4
4	3							Sicherheitsinformation: Wichtige Informationen zu den Sicherheitshin	• Um die Entwicklung eines entflamm		baren Gas-Luftgemisches bei einem Leck des K	일반	5

해당 부분 제품 정보 입력 필요

PDF Parsing 가이드 – 검수 시 참고사항

1. 그림 캡션항목이 다음에 오는 내용에 통합

- 10번 '이미지/표 박스 영역'(보라색) 내부 구조를 투컬럼으로 인식하여 12번 '내용' 다음에 13번 '내용' 순으로 레이아웃 순서를 지정
- 알고리즘상 13번 '내용'의 시작 위치가 12번 '내용'의 시작 위치보다 뒤에 있으므로 해당 두 '내용'은 하나로 통합



참고 이미지

nglichen Stärke behalten möchten.	참고
A. Modelle vom Typ A B. Modelle vom Typ B 3. Aktivieren Sie die Funktion AutoFill-Krug auf dem Bedienfeld. Öffnen Sie die Tür des Getränke-Centers (Typ A) oder die linke Tür des Kühlabteils (Typ B) und setzen Sie den Wasserkrug in die Halterung ein. Schieben Sie ihn, wie in der Abbildung gezeigt, gerade hinein.	
	참고

검수 전/후

nglichen Stärke behalten möchten.	참고
A. Modelle vom Typ A B. Modelle vom Typ B	참고
3. Aktivieren Sie die Funktion AutoFill-Krug auf dem Bedienfeld. Öffnen Sie die Tür des Getränke-Centers (Typ A) oder die linke Tür des Kühlabteils (Typ B) und setzen Sie den Wasserkrug in die Halterung ein. Schieben Sie ihn, wie in der Abbildung gezeigt, gerade hinein.	참고

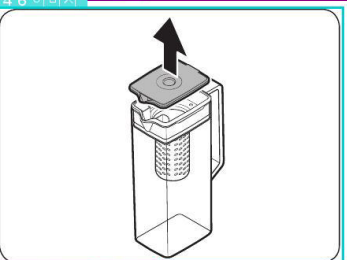
PDF Parsing 가이드 – 검수 시 참고사항

2. '내용' 중간에 경고, 주의, 참고가 있는 경우 뒤의 '내용'들은 해당 'content_tpye'으로 유지됨
- 어디까지가 해당 'content_type'인지 알고리즘으로만 구분하기 어렵기 때문에 다음 '대제목', '중제목', '소제목'이 나올 때까지 해당 'content_type'을 유지
 - 해당 방법이 검수할 'content_type' 수가 더 적음

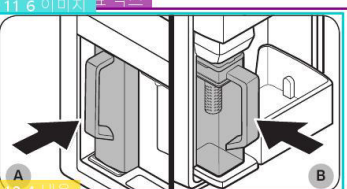
2 3 소제목

So füllen Sie den Wasserkrug

4 6 이미지



11 6 이미지



12 4 내용

A. Modelle vom Typ A
B. Modelle vom Typ B

5 4 내용

1. Fassen Sie in die Griffmulde des Deckels und ziehen Sie den Deckel dann nach oben, um ihn zu öffnen.

2. Geben Sie nach Geschmack Teeblätter oder Früchte in das Teesieb.

6 3 내용

Deaktivieren Sie die Funktion AutoFill-Krug, wenn Sie den aufgebrühten Tee in seiner anfänglichen Stärke behalten möchten.

13 4 내용

3. Aktivieren Sie die Funktion AutoFill-Krug auf dem Bedienfeld. Öffnen Sie die Tür des Getränke-Centers (Typ A) oder die linke Tür des Kühlabteils (Typ B) und setzen Sie den Wasserkrug in die Halterung ein. Schieben Sie ihn, wie in der Abbildung gezeigt, gerade hinein.

14 4 내용

A. Modelle vom Typ A
B. Modelle vom Typ B

15 4 내용

4. Schließen Sie die Tür. Kurze Zeit später wird der Wasserkrug mit gekühltem Wasser gefüllt.

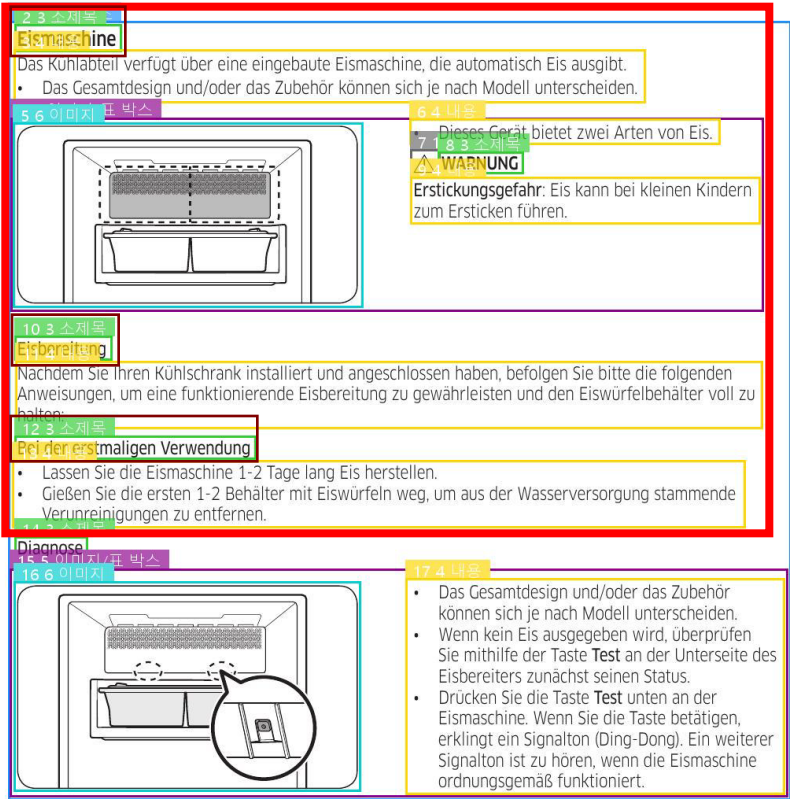
참고 이미지

01 Deckel 02 Teesiebhalter 03 Teesieb 04 Gehäuse	일반	02
1. Fassen Sie in die Griffmulde des Deckels und ziehen Sie den Deckel dann nach oben, um ihn zu öffnen.	일반	63
2. Geben Sie nach Geschmack Teeblätter oder Früchte in das Teesieb.	일반	63
Deaktivieren Sie die Funktion AutoFill-Krug, wenn Sie den aufgebrühten Tee in seiner anfänglichen Stärke behalten möchten.	참고	63
A. Modelle vom Typ A B. Modelle vom Typ B	참고	63
3. Aktivieren Sie die Funktion AutoFill-Krug auf dem Bedienfeld. Öffnen Sie die Tür des Getränke-Centers (Typ A) oder die linke Tür des Kühlabteils (Typ B) und setzen Sie den Wasserkrug gerade hinein.	참고	63
1. Fassen Sie in die Griffmulde des Deckels und ziehen Sie den Deckel dann nach oben, um ihn zu öffnen.	일반	63
2. Geben Sie nach Geschmack Teeblätter oder Früchte in das Teesieb.	일반	63
Deaktivieren Sie die Funktion AutoFill-Krug, wenn Sie den aufgebrühten Tee in seiner anfänglichen Stärke behalten möchten.	참고	63
A. Modelle vom Typ A B. Modelle vom Typ B	일반	63
3. Aktivieren Sie die Funktion AutoFill-Krug auf dem Bedienfeld. Öffnen Sie die Tür des Getränke-Centers (Typ A) oder die linke Tür des Kühlabteils (Typ B) und setzen Sie den Wasserkrug in die Halterung ein. Schieben Sie ihn, wie in der Abbildung gezeigt, gerade hinein.	일반	63
Stellen Sie sicher, dass der Wasserkrug in die Halterung passt.	주의	63
4. Schließen Sie die Tür. Kurze Zeit später wird der Wasserkrug mit gekühltem Wasser gefüllt.	일반	63

검수 전/후

PDF Parsing 가이드 – 검수 시 참고사항

3. '소제목' 다음에 오는 '소제목'이 기존 '소제목'보다 하위 항목인 경우, '내용'에 통합시키고 기존 '소제목'을 유지
- '대제목', '중제목', '소제목' 보다 하위 항목은 클래스에 정의되지 않고, 케이스도 적기 때문에 '내용'에 통합



참고 이미지

Wassertrübung	nach einigen Sekunden wieder klar sein.	일반	66
Eismaschine	Das Kühlteil verfügt über eine eingebaute Eismaschine, die automatisch Eis ausgibt.	일반	67
Eismaschine	• Das Gesamtdesign und/oder das Zubehör können sich je nach Modell unterscheiden.	일반	67
Eismaschine	• Dieses Gerät bietet zwei Arten von Eis.	경고	67
Eismaschine	Erstickengefahr: Eis kann bei kleinen Kindern zum Erstickten führen.	일반	67
Eisbereitung	Nachdem Sie Ihren Kühlschrank installiert und angeschlossen haben, befolgen Sie bitte die folgenden Anweisungen, um eine funktionierende Eisbereitung zu gewährleisten und den Eiskübelbehälter voll zu halten:	일반	67
Bei der erstmaligen Verwendung	• Lassen Sie die Eismaschine 1-2 Tage lang Eis herstellen. • Gießen Sie die ersten 1-2 Behälter mit Eiskübeln weg, um aus der Wasserversorgung stammende Verunreinigungen zu entfernen.	일반	67

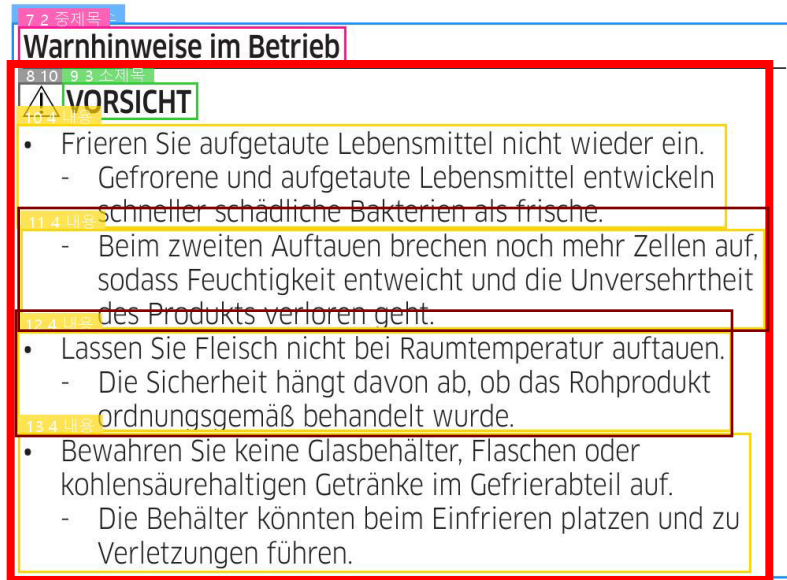
검수 전/후

Wassertrübung	nach einigen Sekunden wieder klar sein.	일반	66
Eismaschine	Das Kühlteil verfügt über eine eingebaute Eismaschine, die automatisch Eis ausgibt.	일반	67
Eismaschine	• Das Gesamtdesign und/oder das Zubehör können sich je nach Modell unterscheiden.	일반	67
Eismaschine	• Dieses Gerät bietet zwei Arten von Eis.	경고	67
Eismaschine	Erstickengefahr: Eis kann bei kleinen Kindern zum Erstickten führen.	일반	67
Eisbereitung	Nachdem Sie Ihren Kühlschrank installiert und angeschlossen haben, befolgen Sie bitte die folgenden Anweisungen, um eine funktionierende Eisbereitung zu gewährleisten und den Eiskübelbehälter voll zu halten:	일반	67
Bei der erstmaligen Verwendung	• Lassen Sie die Eismaschine 1-2 Tage lang Eis herstellen. • Gießen Sie die ersten 1-2 Behälter mit Eiskübeln weg, um aus der Wasserversorgung stammende Verunreinigungen zu entfernen.	일반	67

PDF Parsing 가이드 – 검수 시 참고사항

4. 불릿으로 시작하는 '내용'은 이전 '내용'에 통합

- 알고리즘으로 구현되어 있지만, 이전 '내용'의 글의 시작 위치가 더 뒤에 있는 경우는 제외하였음
- 11번 '내용'의 시작 위치가 12번 '내용'의 시작 위치보다 더 뒤에 있으므로 해당 '내용'은 알고리즘에서 통합되지 않음



참고 이미지

• Frieren Sie aufgetaute Lebensmittel nicht wieder ein. - Gefrorene und aufgetaute Lebensmittel entwickeln schneller schädliche Bakterien als frische. - Beim zweiten Auftauen brechen noch mehr Zellen auf, sodass Feuchtigkeit entweicht und die Unversehrtheit des Produkts verloren geht.		주의	24
• Lassen Sie Fleisch nicht bei Raumtemperatur auftauen. - Die Sicherheit hängt davon ab, ob das Rohprodukt ordnungsgemäß behandelt wurde. • Bewahren Sie keine Glasbehälter, Flaschen oder kohlenensäurehaltigen Getränke im Gefrierabteil auf. - Die Behälter könnten beim Einfrieren platzen und zu Verletzungen führen.		주의	24

검수 전/후

• Frieren Sie aufgetaute Lebensmittel nicht wieder ein. - Gefrorene und aufgetaute Lebensmittel entwickeln schneller schädliche Bakterien als frische. - Beim zweiten Auftauen brechen noch mehr Zellen auf, sodass Feuchtigkeit entweicht und die Unversehrtheit des Produkts verloren geht. • Lassen Sie Fleisch nicht bei Raumtemperatur auftauen. - Die Sicherheit hängt davon ab, ob das Rohprodukt ordnungsgemäß behandelt wurde. • Bewahren Sie keine Glasbehälter, Flaschen oder kohlenensäurehaltigen Getränke im Gefrierabteil auf. - Die Behälter könnten beim Einfrieren platzen und zu Verletzungen führen.		주의	24
--	--	----	----